



# **HOMOLOGOUS PROTEINS:**

## Top BLAST Hits:

	Score	E
gb AAF81911.1 AF279265_1 (AF279265) putative anion transpor...	476	e-133
gb AAF71715.1 AF230376_1 (AF230376) prestin [Meriones ungui...	471	e-131
ref NP_000432.1  pendrin [Homo sapiens] >gi 11421915 ref XP...	451	e-125
ref NP_035997.1  Pendred syndrome homolog (human); Pendred'...	448	e-124
ref NP_062087.1  Pendred syndrome homolog (human) [Rattus n...	447	e-124
ref NP_067328.1  down-regulated in adenoma [Mus musculus] >...	434	e-120
ref NP_000102.1  down-regulated in adenoma protein [Homo sa...	418	e-115
sp O70531 DTD_RAT SULFATE TRANSPORTER (DIASTROPHIC DYSPLASI...	365	1e-99
ref NP_000103.1  sulfate anion transporter 1; Diastrophic d...	362	1e-98
ref NP_031911.1  diastrophic dysplasia [Mus musculus] >gi 2...	357	4e-97

## BLAST to dbEST:

	Score	E
gi 8630793 /dataset=dbest /taxon=960...	523	e-146

## **EXPRESSION INFORMATION FOR MODULATORY USE:**

library source:

### Expression information from BLAST dbEST hits:

gi|8630793 Human head-neck

### Expression information from PCR-based tissue screening panels:

Human fetal lung

1 MSQPRPRYVV DRAAYSLTLF DDEFKKDRT YPVGEKL RNA FRCSSAKIKA  
 51 VVFGLLPVLS WLPKYKIKDY IIPDLLGGLS GGSIQVPQGM AFALLANLPA  
 101 VNGLYSSFFP LLTYFFLGGV HQMVPGTFAV ISILVGNICL QLAPESKFQV  
 151 FNNATNESYV DTAAMEAERL HVSATLACLT AIIQMGLGFM QFGFVAIYLS  
 201 ESFIRGFM TA AGLQILISVL KYIFGLTIPS YTGPGSIVFT FIDICKNLPH  
 251 TNIASLIFAL ISGAFLVLVK ELNARYMHKI RFPIPTMIV VVVATAISGG  
 301 CKMPKKYHMQ IVGEIQRGFP TPVSPVVSQW KDMIGTAFSL AIVSYVINLA  
 351 MGRTLANKHG YDVD SNQEMI ALGCSNFFGS FFKIHVICCA LSVTLAVDGA  
 401 GGKSQVASLC VSLVVMITML VLGIIYLYPLP KSVLGALIAV NLKNSLKQLT  
 451 DPYYLWRKSK LDCCIWVVSF LSSFFLSLPY GVAVGVAFSV LVVVFQTQFR  
 501 NGYALAQVMD TDIYVNP KTY NRAQDIQGIK IITYCSPLYF ANSEIFRQKV  
 551 IAKTVSLQEL QQDFENAPPT DPNNNQTPAN GTSVSYITFS PDSSSPAQSE  
 601 PPASAEAPGE PSDMLASVPP FVTFTLILD MSGVSFVDLM GIKALAKLSS  
 651 TYGKIGVKVF LVNIHAQVYN DISHGGVFED GSLECKHVFP SIHDAVLFAQ  
 701 ANARDVTPGH NFQGAPGDAE LSLYDSEEDI RSYWDLEQEM FGSMFHAETL  
 751 TAL

**FEATURES:**

**Functional domains and key regions:**

[1] PDOC00001 PS00001 ASN\_GLYCOSYLATION  
N-glycosylation site

Number of matches: 3

1 153-156 NATN  
 2 156-159 NESY  
 3 580-583 NGTS

[2] PDOC00005 PS00005 PKC\_PHOSPHO\_SITE  
Protein kinase C phosphorylation site

Number of matches: 2

1 45-47 SAK  
 2 445-447 SLK

[3] PDOC00006 PS00006 CK2\_PHOSPHO\_SITE  
Casein kinase II phosphorylation site

Number of matches: 11

1 18-21 TLFD  
 2 158-161 SYVD  
 3 240-243 TFID  
 4 365-368 SNQE  
 5 459-462 SKLD  
 6 556-559 SLQE  
 7 635-638 SFVD  
 8 691-694 SIHD  
 9 722-725 SLYD  
 10 726-729 SEED  
 11 732-735 SYWD

[4] PDOC00007 PS00007 TYR\_PHOSPHO\_SITE  
Tyrosine kinase phosphorylation site

Number of matches: 2

1 7-15 RYVVDRAAY  
 2 447-454 KQLTDPYY

003321" 68564260

[illegible]

Number of matches: 10

1	77-82	GGLSGG
2	78-83	GLSGGS
3	89-94	GMAFAL
4	103-108	GLYSSF
5	335-340	GTAFLS
6	435-440	GALIAV
7	481-486	GVAVG
8	485-490	GVAFS
9	581-586	GTSVS
10	681-686	GSLECK

Membrane spanning structure and domains:

Helix	Begin	End	Score	Certainty
1	51	71	0.893	Putative
2	82	102	1.020	Certain
3	107	127	1.729	Certain
4	130	150	1.497	Certain
5	186	206	1.723	Certain
6	228	248	1.517	Certain
7	256	276	1.898	Certain
8	288	308	1.252	Certain
9	338	358	1.568	Certain
10	383	403	1.304	Certain
11	412	432	2.345	Certain
12	469	489	1.997	Certain
13	619	639	1.146	Certain

**097368-1**

Length = 738

Query: 54 LFDDEFEKEDR--TYPVGEKLRNAFRCSAKIKAVVFGLLPVLWSLPHYKIKDYIIPDLL 227  
L + EE R + P + R +CS A+ A++ LPVL WLP+Y ++D+++ DLL  
Sbjct: 15 LNQEHLEELGRWGSAPRTHQWRTWLQCSRARAYALLLQHLPLVVLPRYPVRDWLLGDDL 74

Query: 408 N I C L Q L A P E S K F Q V F N N A T N E S Y V D T A A M E A E R L H V S A T L A C L T A I I Q M G L G F M Q F G F V A 587  
 ++ LAP+ A N+S ++ A +A R+ V++TL+ L + Q+GLG + FGFV  
 Sbict: 135 S V T E S L A P O-----A L N D S M I N E T A R D A A R V O V A S T L S V L V G L F Q V G L G L I H F G F V V 186

Query: 768 IFALISGAFLVLVKELNARYMHKIRFPIPTEMIVVVVATAISGGCKMPKKYHMQIVGEIQ 947  
+ A ++G LV+VK LN + ++ PIP E++ ++ AT IS G + ++ + +VG I  
Sbjct: 247 VTAAGVGVVLVVVKLLNDKLOOQLPMPPIGELLTLIGATGISYGMGLKHRFEVDVVGNI 306

Query: 1128 FFGSFFKIHVICCALSVTLAVDAGGKSKQVASLCVSLVVMITMLVLGIYLYPLPKSVLGA 1307  
 G F+ + C++S +L + GG SQVA SL +++ ++ LG + LPK+VL A  
 Subject: 367 LIGGIFOCFPVSCSMRSLSVOESTGGNSQVAGAISSLFILLIIVKLGELFHDLPKAVLAA 426

Query: 1488 TQFRNGYALAQVMDTDIYVNPkTYNRAQDIQIGIKIITYCSPLFYANSEIF----- 1637  
TQ + L QV DTDIY + Y+ A++++G+K+ + +YFAN+E +  
Sbjct: 487 TOMPHYSVLGOVPDTDIYRDVAEYSEAKEVRGKVKFRSSATVYFANAEFYSDALKORCGV 546

Query: 1761 -----TFSPDSSSPAQSEPPASAEAPGEPDMLASVPPFVTFHTLILDMSGVSFVDLMGI 1925  
S D A + ++AP + S + A P FH+LILD+ +SFVD + +  
Sbjct: 607 CKMMOVSSGDKMEDATANGOEDSKAP-DGSTLKALGLPOPDFHSLILDGLALSFVDTVCL 665

Query: 1926 KALAKLSSTYGKIGVKVFLVNIHAQVYNDISHGGVFEDGSLECKHVFPSIHDAVLFAQAN 2105  
K+L + + +I V+V++ H+ V + + G F D S+ KH+F S+HDAV FA +  
Sbjct: 666 KSLKNIFHDFREIEVEVYMAACHSPVVSOLEAGHFF-DASITKKHLFASVHDAVTALQH 724

Query: 2106 ARDV 2117  
R V  
Sbjct: 725 PRPV 728

Hammer search results (Pfam):

Model	Description	Score	E-value	N
PF00916	Sulfate transporter family	254.5	1.5e-72	1
PF00189	Ribosomal protein S3, C-terminal domain.	3.3	8	1

Parsed for domains:

Model	Domain	seq-f	seq-t	hmm-f	hmm-t	score	E-value
PF00916	1/1	187	497 ..	1	328 []	254.5	1.5e-72
PF00189	1/1	651	661 ..	79	89 .]	3.3	8



3151	AACTAACAAAT	TCCTATGATA	CTGGCATGAG	AGCCTTGAAC	AGTGCGCTGGC
3201	ATAGAGAAGG	TGCACCAATA	AATATTTGTT	TCATGAATGA	ATGAATGAAT
3251	GAATGTCTAG	AAAGCTAATC	CCTCTCAGCC	TCTGTTTCCA	GTTCTTCTTT
3301	CAAGCTTCAG	ATTGCTTTGC	CCAACATACA	GCAGACTTGC	AAGTAAGGTT
3351	GGGCATGGAC	TAGCCCTCAA	ATGAGTTGTT	TTTCTTTCCC	TAGCCAGCTC
3401	TCTATTACATA	AGTCGGGCTT	TCTCTGCCAC	AAACAGACCT	GATGGAGGCC
3451	CTGCAGGGCT	GGTTCTCTCT	TCAAGCAAGG	CTTTAGAGTT	GCATTTAAGCA
3501	ATTTATCCCC	CGTCCACCTC	CCCTTCCAGC	ATCCCAGGGA	TGGCAGAGGC
3551	ACCCATGAGC	CCCAGAAGGG	ACAGGGGGTA	AGATATTGAT	GATGATGCTT
3601	TTTCTTGGAG	TGTTAGTTGG	AAGAGAAAAT	CTGCCCAGAC	TTTCCAAGGT
3651	ACAAAGCATT	GTCTTTGTTG	GTTTCACTCT	TGGGTGACAT	CCAGGGGACC
3701	GAGTGTCCAG	GAACATATTT	TTGAGCAAGA	GCAAAGAGCA	GGAAATGGTG
3751	CTGGGCAGGA	AAGGAAGCCT	CATCAGAGCA	GGCCAGTGAG	TACCAAAATG
3801	GGCCCTAAGT	ATTTGAGTTC	CCTCAACTGG	GAGAAGGAAA	GCAAAATGCC
3851	CTCACCCACT	TCCAGTCATC	AATCCACCGG	CTGTCACCCT	TGAGTTTGTA
3901	AGCCCTTGTT	CCTACCGCTC	CTGAGTTTCT	ATGAAAGGAC	CTTGAGGTGT
3951	TCAACAAAACA	GGGAAGGGAT	CAACTCTCCC	CACCTTGCGT	TGACCAATGA
4001	ATTCTTCCCT	CCTCTGCTGC	CCAGTGAATT	AACAGGAGAA	AGAACTCCGG
4051	TATTTGAGTT	ACCACATATC	AAGTAGATAG	AGTCAGCAGA	GTGCACCCTG
4101	CAGGACAACAT	AGAGCCTTCC	TTTTCAAGGA	AGTTCTAAGA	AAAAATGGCAG
4151	CAGGCAGGCC	CCACTCGGGT	GTATTCACTC	ATTCAATTTAT	TCAACAAATA
4201	TTTACTAAGT	GGCCCTGTGC	AAGGCTCGAG	GTGTACAAAG	ATGAACAGGA
4251	GAGCTAGACT	TCCTGCCATG	CGTGGTGGGG	TTTGCTGCCT	AGTGGGAGAG
4301	ACAGACAAAA	AGCAAGGAAT	GCACACACAG	GATGCACAGA	CAGCGGCAGG
4351	AACCAAGGTG	CAGTTAACCA	GGCCTGGGAT	CAGACAGACA	GGACTCAGAG
4401	GAGACTTTCC	CAGAGAAAAG	CCATCTGAGC	CAAGGATGAT	ATCTGATACC
4451	TCCGAAGGCT	GAGCCACCAT	AACACTCATA	CCTTTAAGCC	AAGTCTTATA
4501	AACTCCCCAG	GTAAGCAGCT	GGCAGTCAGA	AGACCTCCAG	CTAATGCCCA
4551	GGACAAGTTG	ATGAGCTCTC	AAGAAAAAGT	TCCTGCCTTT	TCCTCTCAAT
4601	ATCCCTGGCA	CACAGTTTCA	TGAATTTTGA	ATGAACCAAT	GAATGAAATG
4651	AGCAGGATAT	GATAATCCCT	CTCCAACACG	GAATGTCCAA	GCCATGCAGA
4701	CGCGACTTGA	AATTTTCCCC	GTTCCCTTCC	AGATGTCTCT	CAGCCAAGAT
4751	CAAAGCTGTG	GTGTTTGGGC	TGCTGCCTGT	GCTCTCCTGG	CTCCCCAAGT
4801	ACAAGATTAA	AGACTACATC	ATTCTTGACC	TGCTCGGTGG	ACTCAGCGGG
4851	GGATCCATCC	AGGTCCACAA	AGGTGAAGGG	GCTCCTTCAG	CCAGGCCTGG
4901	ATTGCCACTC	CCCTCACCAT	TCCTCTCCTC	ATCCCCACTC	CATCCCTCTG
4951	TGATCCCCAT	AAGCTAGTCA	TGCTGCTGAG	CTTCACTCTC	GTTGTCTCTT
5001	CGAGGCATGG	CATTTGCTCT	GCTGGCCAAC	CTTCTGTCAG	TCAATTGGCT
5051	CTACTCCTCC	TTCTTCCCCC	TCCTGACCTA	CTTCTTCTCT	GGGGGTGTTT
5101	ACCAGATGGT	GCCAGGTAAG	GCCTCTCCCC	TCTGGGCAGG	CAGGATGACC
5151	CAGACCACAA	GGATGGGAGG	TGTGGCAAAG	GGGCCTCGGG	AGATTTTCCA
5201	TCTGCATTCT	CCTGGAGTTG	TTCTGGTCA	GTCTAGGGG	AATGGTCACT
5251	GTGAATGTCA	TTTCCAGGTC	CTCGGTGACC	TTGGAGAAAC	CACCTGAGCT
5301	CTTTGAGTTC	AGTTAGCATT	ACCTGTTCCA	TCTTCTCCTT	AGGAATGAGA
5351	GGAAGACTTA	GCAGAAACAAG	ATATACCATA	TGCTATAACA	TGCTTAAACA
5401	GATGTGAGAA	ATCACCATCT	AACTCCCTGG	TTGGTCCCAG	CCGGCCACTA
5451	CAGGGACATT	TGGACTTCTC	TGGTGCTAAG	TGAGATGGAG	GAAAGCCTGG
5501	TCACAAGGGC	TGGTTTCTGG	TTCAGGCTCT	GCTTATATTT	CTTATTTCTG
5551	AGTTCAATTT	CTCACGTGTC	CTGTATGACA	ATATTGACCA	TTGGGGTAAA
5601	AGCACTTGA	AAAGCATAGA	TCATGGTTAG	AGTGAGTGGT	TGTTATTATT
5651	TGTTTGGAGA	AGAGCCTTGG	AGGTGCAGGG	ATCATATCCC	CTGGGATCGG
5701	GAAGCATTCC	TGGGCCCCCT	TCTGGTTTCC	ATCGGTGTGG	TTCAAACCTC
5751	TGATTTTTGC	TGGCTGGGTG	GGGCACCACA	GGTACCTTTG	CCGTTATCAG
5801	CATCCTGGTG	GGTAACATCT	GTCTGCAGCT	GGCCCCAGAG	TCGAAATTCC
5851	AGGTCTTCAA	CAATGCCACC	AATGAGAGCT	ATGTGGACAC	AGCAGCCATG
5901	GAGGCTGAGA	GGGTGCACGT	GTGAGTACG	CTAGCCTGCC	TCACTGCCAT
5951	CATCCAGGTG	AGGGGCGCAG	CCCCAACCTT	GCTAGAAGGG	CATCAGACCA
6001	CCCTGCCCCC	CCCTCAAAGC	CTTAGCTTTG	ATGCTAAATC	TGATTTAGGG
6051	GGCTGGGTGT	GGAGGCTCAT	GCCTGTAATC	CCAGCACTTT	GGGAGGCTGA
6101	GGAGGTGGA	TCACTTGAGG	TCAGGAGTTT	GAGACCACCT	TGACCAACGT
6151	GATGAAACCC	CATCTCTACC	AAAATATCAA	AAATAATCCA	GGCTTGGTAG
6201	TATGCGCCTG	TAGTCCCACC	TACTCAGGAG	TGTGAGGCAG	GAGAATCACT
6251	TGAATCCGGG	AGGCAGAGGT	TGCAGTGAGC	GAGATTCGCG	CCACTGCACAT



6301	CCAGCCTGGG	TGACAGAGCG	AGACTCCGTC	TCAAAAAAAAA	AAAAAAAAA
6351	AAAAAAAAA	CCCAAGTTAG	GGCTCACCTC	CTCCCTCCTC	CCCATCCCAG
6401	GGCTAAAGTG	AACCTTGAAA	ATTAACAGTA	TCTCCTCATC	TGCATGTAGC
6451	AGGACCATAC	AAAAAAAAACA	CAGCTGTACC	TGGTTAAACT	GTCTTGAGCT
6501	TTAAACCTGT	AAAAGACTCA	CAGCCTCTCT	CCATTATCCC	GTGGAGAAAC
6551	CCAACCTCTC	GCCAGCATAG	TCTTGCGACG	TGCTTAATTTT	CTCTAACATC
6601	CCTCACTCCG	CTCCAGCCTC	CTCTGCTCCA	AGCCACAGCA	GCAGTTGCAC
6651	AACATAAATT	GAGCTTCTGC	AAATGGTTGC	AAAGGATTCT	GCTAGGTTTT
6701	ATGAAGGGAA	GCACAACATG	ACAGAATGCA	AGAGCAAAC	ACAGTCCCAG
6751	AGAGCGCCTT	TTCATTCACT	CATTCAATCG	GTTTTGTGCC	AAGAACTAGG
6801	CTAAACCTCG	GGATACAAAG	ATAAGTAAGA	AAGAGGTCCA	ATTCACAAGT
6851	TGCTCACAGC	CCAGCAGAGG	AAGGAGCCAT	GTCAACAGAT	AAATTTGTAT
6901	GCAGTGAGAT	AAGCAGCAAA	GTAGAGCCAT	GTACAAAGAC	TGTTAGGGACA
6951	CAGAGCAGAG	TCACGGAGGA	CCTCAAAGAG	GAGGTGACAC	TCCACCTCTC
7001	TTAAAGGATG	AGAACTTAAC	CAGGAACAAG	GTATACAGAG	GATGGTCCAG
7051	GCAGAAGGGA	ACAGTGCTTA	AAAACACTGA	GGCCTGAGAG	AGTGTGATCT
7101	GCGCAGGCCA	AGTAAGGGGC	TTGGTGTGGC	TGGAGGGTAG	AGGGCCCCAG
7151	AGAGAGTAGG	AAAGTAGGCA	CGAGCCAGAC	AATGAGATCT	GGGCTCTGTT
7201	CTCTGCACGC	GACTTTGGGT	CTGATTTGGCA	TTTATAAAGG	ATCGTTTGGG
7251	CTACACAATG	ATGAGTGCGA	GGTGGATTAG	AATCAAGGCA	GGGGACCTGT
7301	TGGGAGACTC	TGCAGAGGCC	CAGGCAGGAA	TAATGCAGGC	GAAGACCAGG
7351	TAGAGAAAGA	GATGGGGCTG	GACTTGAAAA	GAATGTTTTA	CCAGGAGCTT
7401	GGTGATAGAC	TGGATGTGGG	AGGTAAGGGA	GGATGACTCT	CAAGTTTTTG
7451	GTTHGGCAAC	CAGGTTAATG	ATGGTGTCA	TTACTGAGAG	AGAAAACTT
7501	GGGGGAGGAC	TAGACTTATT	TTACAGATAA	GCCAAAGCCA	GAGAGGTGAT
7551	GTGACAGAAA	GGCCCATGCT	CTAAAGGAGC	TGAAGGTCTG	ATGGCAGCCA
7601	TGTAGAGCAC	AGTGAAGGGC	AGGTGAAGGT	CACAGATGGT	CAAATTCCTT
7651	CAAGCTACTG	CTACGCTAGG	ACTGCACGGA	GCTCCAGACC	TGCGTGTGTG
7701	TGGGGCGGGT	CGTTGGAACT	GCTGAACCAC	ATTGCTCTTC	CGCCACCAAC
7751	CACCCTTTTT	CTCCTCTCAG	ATGGGTCTGG	GCTTCATGCA	GTTTGGCTTT
7801	GTGGCCATCT	ACCTCTCCGA	GTCTTCTATC	CGGGCTTCCA	TGACGGCCCG
7851	CGGCTGTCAG	ATCCTGATT	CGGTGCTCAA	GTACATCTTC	GGACTGACCA
7901	TCCCCTCTTA	CACAGGCCCA	GGGTCCATCG	TCTTTGTGAG	TCTGGGGATG
7951	CACCCCTGCC	ATTGGAGCAA	GGCTCCAGCA	GACACATGAG	GAGGATGTAC
8001	TGTTTTTAAGA	TGTCGTGAGC	TCCTCATTGC	AAGGGCTGGC	TTAGCTGTTG
8051	TTCAGAGAGG	ATTCTGAGGG	GGTTTCTGTC	TTGGGAGGGT	CAAAGTCATG
8101	ACTCACAGAG	GTTCCTGGTA	GTTAATACCT	CGAGAAAAGT	GCTGTACATT
8151	CTCCGCGCAGT	TCCCCATTCT	AGTGCTCTCA	CCCCCTCCCTG	CCTGGAAAGT
8201	CCTGCCTTAT	GTCTAATCTC	CATCCCTCCT	CCTTCAGCCC	AAACTCTTCT
8251	AAAGAAAAAG	AAAGCATTCC	TTTTCTAGCA	CAAGTTCCCC	ATGTGCCTTT
8301	TGGGAAAGGG	CGGTGGGCGA	CGGGACAGGG	TTCTGATCA	GGGTTTTAAT
8351	TCTGTCTTGG	TGTGCCTCCA	TTAGCTTTGA	TGGCATCCCT	TCCCTGGGTC
8401	AGACACCCAA	AGGTGGGGTA	TTATGGGAAG	AAGGGTGGG	AGCCTGTGAG
8451	CATGATGCTC	TTTTCCCCAG	ACCTTCATTG	ACATTTGCAA	AAACTGCCCC
8501	CACACCAACA	TCGCCTCGCT	CATCTTCGCT	CTCATCAGCG	GTGCCTTCCT
8551	GGTGTCTGGT	AAGGAGCTCA	ATGCTCGCTA	CATGCACAAG	ATTGCTTCC
8601	CCATCCCTAC	AGAGATGATT	GTGGTAAGGA	CCTTGTTTCA	AGCTGGGATG
8651	TTGGGGGGGC	AGGCTGTGAG	ACGAGGAAGC	CCCTACCTTT	CCTCACCCCA
8701	TCCCCCTCAAC	TGGCAGCCAG	TGGGACAGGA	AGTCAGTTGT	GAATCCATCC
8751	CATCCCCGCT	ATGTGCGGTT	TCCTTCTTTT	CTACTGCTCT	ATAAATTCCC
8801	CCTAAGGAGG	CAGGGGAGTG	GGATTCTAGG	TCCCCAGAGA	AAAGGGAGAC
8851	TTGAGAGAGA	CGCCTGCCCT	GGCCCCACCT	TAGGGCCAAT	CCCCATTCTC
8901	CACTCTGGGG	TTTGCAGGTG	GTGGTGGCAA	CAGCTATCTC	CGGGGGCTGT
8951	AAGATGCCCA	AAAAGTATCA	CATGCAGATC	GTGGGAGAAA	TCCAACGCGG
9001	GTGAGTCCAG	GTGGCCCCAG	AGCCTGGCCG	ACCCGCACCT	CATGCCCCAC
9051	TAAAGCCTGA	GCTCGGAGAG	GGAGACAAGA	TGAACCTCAT	GAAAGTGCAG
9101	TCGAAACTGT	ATGACACTGA	CCATGTATGA	ATTATTACTA	TTACCGTTTC
9151	CTGAGAAGGG	CCGCACAACC	AGCCAATGTA	GGCTATTTTA	TGAGAAATGA
9201	GTCTTAACTG	CCACACTCCC	CTTATAAATC	TCATTCAACT	GATGCTGTTA
9251	AACAAAGCCT	CTCTGAACAG	CCGCTTGCTG	GCTCTTGCC	TTGTCTAAT
9301	GCATTGGTTC	TTTGTCCTAG	TAGAAAGGGA	ACTATTAGGT	TCAACCAGAT
9351	TCATGAAGCA	TCCACTCTGT	GCCAGGCACC	ATGCTGGGCC	CTGGGAGGAG
9401	AGGGGTGACG	CTTGTCTCTG	AGGGTTGGAA	CAGGCAAGGG	AGGGAAGACC

```

9451 ACATAGCACC AAAGGTCTAG GGGTCTGTGG ACTCGTGAGC ATACAGGGTT
9501 CAGAATCTGG GAGTTAACAA ACGAGGCCCT ACCACATACT GGCCCCGGGA
9551 CCTTGGGCAA GTTAGGTTCT CTCAGCCTCA GTTTCCTCCT TTGTAAACA
9601 GGAGTAGTGG TCCCTACCTT ATGGGGTGGT GCTGAGGATT CAGACTGGAT
9651 GGGATAACTT AGGCAAAGAT CCCGGCACAC CATGGGGGCC TGGCTGGTCC
9701 CTGTGGGCTG GTGAAGGACT TGGCTGCCCT CCCCCTCAC ACCCTTGGGT
9751 TCTGCCTCCT TCCTGGCTCC TCGGCAGGTT CCCCACCCG GTGTCGCCTG
9801 TGGTCTCACA GTGGAAGGAC ATGATAGGCA CAGCCTTCTC CCTAGCCATC
9851 GTGAGCTACG TCATCAACCT GGCTATGGGC CGGACCCTGG CCAACAAGCA
9901 CGGCTACGAC GTGGATTGCA ACCAGGTAGC TCTGGCCACC CCCGGCAGGA
9951 CTGGGCAGGA CAGGTCAACT CAGGCCGTGGC ATGACATATC TTGGGTGGGG
10001 AGATCATTTG GCTGAGGTGA GGCAGGCTGC CTCGAGTGTG GGGGATAGGG
10051 GGTCTCTGTA CCCTAAGAGG CTGACCTCCT CTTGACTGGG AATGTGTGAC
10101 TTTATAGCCA CTGGGTCACT CTCAGGTCTT AGGCCACAG TCCAGCTTGC
10151 ATGCCTGACT GCACTTGGTC CCCGTGCCCC CCAGCCCCAC ACTGGCTTCT
10201 AATCCTGTCC CCTCCCTGCA GGAGATGATC GCTCTCGGCT GCAGCAACTT
10251 CTTTGGCTCC TTCTTTAAAA TTCATGTCAT TTGCTGTGCG CTTTCTGTCA
10301 CTCTGGCTGG GGATGGAGCT GGAGGAAAAT CCCAGGTGAG CTTTGTCTTA
10351 GGGGAGTTGG GGGGAGGTGG TAAGAGAACA GTTGCCCCAA AAAAGCCTGG
10401 GCACTGCAAG CCAGGCCAGC TCTTCTCCGA CCCCTTCTTC CCGTACTTAG
10451 TCTCCACTCC ACCAAAGCCA TGGATTGGAA ATAAATCAAG AGCAAAAATT
10501 TCACACCTTC CCTCTATCCC CAACTCTTTC TCGGAATAGG TGGCCAGCCT
10551 GTGTGTGTCT CTGGTGGTGA TGATCACCAT GCTGGTCTCG GGGATCTATC
10601 TGTATCCTCT CCCTAAGGTA AGAGCCCAGC CATCGAGCAG AAGTCAACGA
10651 AAGACTCCAA TAAGAACAAT CCCTGAGAGT TGTGTGGCAC TTTACGGACC
10701 ACAAAGTGCC ACTGTTGTCA TACTTAGTCT CAACCACAAA CTGTGAGGTA
10751 GACAATGCAG GTTTTATCCT CCCCATTTTA CAGGTGAAGG AAAGTGTGTC
10801 TGAGAGTCTA AGTAACCTTG TCCATAGTGA GGCAGCTTAC AGCGCAGGGC
10851 TGGTCCCAAA CTCCAGCCTT CTGGCCTCAG AGTCTAATCC CTAGGCAACA
10901 TTTGCACCTA CCCACGAGTA CCAGGCTCTT ATATAGCCCA GCTAGGAGGG
10951 CTCTAGGCAT CGCTCATTTA GAGATGAGGG AAGAGAGATA GGGAAAGGAT
11001 GGGGCCAGGA AGGACCCCAT GGCTCTAACG CCAGCACTTT CCAAACCTAA
11051 GGTCGAATGC AGAGATTTGG GGGATCAGCC AGGGGAGGTG TTCCAGAACT
11101 CCGTCTCTGT CCTGCCAGGC CTTGGGGTCG GGTATGCGCA GGAGGGCAAA
11151 AAGAAGGGGA GACCCTGGGG TCCTGGAGCA ATGTTCTGCT TCTCTAGTCT
11201 GTGCTAGGAG CCCTGATCGC TGTCAATCTC AAGAACTCCC TCAAGCAACT
11251 CACCGACCCC TACTACCTGT GGAGGAAGAG CAAGCTGGAC TGTGTAAAGTA
11301 TCGGGCAGCC TCTGGGTACT GGCCATGCCC CTGCCCTCTC CTCCAACCCC
11351 ACAGCCCTGT CAGCCCTGTC CTAACAATGA ACCCTCTAGT CTGCTGCTTC
11401 CTAATTAGCA TGAGATGAGT GGTTAAAAGT CCGAGTTTCG AAGTGAAACA
11451 TCCTATGTTC AAACCCTAAC TCAGCCATCT GCTGGCTCCA TGGCCAATAG
11501 CAAGCCCCCTT AACCTTTCCC AGTCTTGGTG TCTTAACTGG GCAAATGGTT
11551 ATTTTATGCT CTCTGCCTCC CAGGGTTTTT TATGAAGAAG AAGCAAGGTA
11601 ATACAAGTAA ACATGTTGTC TACATCGTAT TTTATACTCA ATAAAGCTTA
11651 GCTATGACTA CTTTATGACA TACAGCTTTA AAAAACAAAA GGAAATAGTT
11701 TGTATTTTAA AAAAAACCT AGAACATAAA GCCAGAGGAC CAAAATCTTG
11751 AGCAAGTTAC TAGACTTCCC TGGGGTTCTA TTTCTCATC TGTAATGGG
11801 GGTGAGACTC ATGCAGTCAT GGTTGCGTCA AACGCTGGTT CCGAGGATTA
11851 AATGAGATCC CAGTGGGAAA ACACCGCATG AGCGCAACA TTCTGCAAAC
11901 ATGACTTATT GTCCTGATTA GTCACACACT CCACCGCATC ATCCGCTGGG
11951 CATAGTAATG AAGCCAGTG TGTTTTGACG ACACTGCCTT CTCTCCATTT
12001 AAGCCCCACC ATAACCTATG GGAGAGGATT TACTAACTT TCTTAACGGT
12051 GATGAAACCA AGGCTCAGAA TGGTTAAGTA AATTGTCAAA GGCCACAGAG
12101 GTAGGGAGTG GTAGAGTCTG GATTAAAAC CCAAGTCTCG GACTCCAGAC
12151 CTCTAGGCTG TACTGTCTCA TAGGGAAGGC AGTCTCAGCC ACCTAGGGCA
12201 GAGAAGAAAA TCCTTAAAGC CAGAGAAGTG AGTGGCTCAT CTGTGGTCAC
12251 CCAGAGAGAC AGTAGTAGG ACAGGGAGAA AAATTATACC TCAGTTCCCA
12301 GCCCAAGGAT CTGCTTTGAC CATAACCCAA CAAGCCCCCG CTATGGTGGT
12351 ATTTCTTAGT GTTCATATGG CGGCTTTTGT TTCCATTTGA TCTTCACAGC
12401 AATTCTCTAC AGGAATCTGG GCAGATTTAT TTCCTTTAGA GGAATTTCCA
12451 GGTCTTAAAA TCTATAGGGG GCAACTATCA AAAGTTTACC CAATGTTGCC
12501 CCCTACCCAC ACACAAAACC AGGCCCCCAG CCGATCAGAA AGCACTGCTG
12551 AGCTCCTGTC AGGGCCCCACG CAGCTCGCTG TGAGACAGAG AGAGGGAAC

```

FIGURE 3, page 4 of 20

12601 CACATTTATT GATCACCTAC TGAGCATCCA TCACTAGGCT AGGACCGTCA  
12651 CATTTCCTTAA CTTTTGAATC CTTTCATGAG GTAGGCATTA TTATTCTCCT  
12701 TTTGTTTCAC ATAGCCATTA AAGAACAAAA TTTGGGGCTG GGTGTGCTGA  
12751 CTCACACCTG TGATCTAGCA CTTTAGGGGG CTGAGGCAGG AGGATCGCTT  
12801 GAAGTCAAGA TTTCAAGGTC AGCTTGCGCA GCTTAGCGAG AGCCGTCTCT  
12851 AGAAAAATAT AAAAGTTAGC TGGGTGTGGT GGCACGTGCC TATAGTCTTA  
12901 ACTATTCAGG AAGGTTAGGC GGGAGCACAA CTTGGGTTC AGGGTTTGAG  
12951 GCTCCAGTGA GCTGATCTTG CCACTGCACT ACAGCCTGAG CAACAGAGCA  
13001 AGACCCCTGT ACTCCAAAA CAAACAAACA AACACATTTT GAACCCAAAC  
13051 AGATCTGACC CAAGATGCAT GCTCTTATAG ATGCCACCTC CTTGTGTGCT  
13101 GGGGCTTCTA CTAACAAACAG AGACAAGATC AGGCAACCAC AGTCAATCTA  
13151 AGGGAAAGAG GAAAGTGTAA CCAAAGCACA AATACATAAA ATATTGCAAA  
13201 AATGCTATTT AAAGAAAAAA AAGAGAAGAG AGGCTCTGAG GTTGTACTAA  
13251 CAGAGAATGG CTTGGCTAA TCCAGGAAGA CTTCTGAAA GAGGTGTGTT  
13301 TTTCCCAGG TCTGCTTTTG ACATCTCTCT TTTACAGTG CATCTGGGTA  
13351 GTGAGCTTCC TCTCCTCCTT CTTCTCAGC CTGCCCTATG GTGTGGCAGT  
13401 GGGTGTGCGC TTCTCCGTCC TGGTCTGTTT CTTCCAGACT CAGTTGTAAG  
13451 TGATAGCTTC CGCCCTCCTA GGCCCAAGT CGGTTCCCTG GGCCAGCCCG  
13501 CAAAGGGCTT CCATGCCACG GCCTGGCTTA GTCCACTGTA CTTCCACCT  
13551 CTGGGCTTGG CACTGGAGGT GCTGCCAGGC CCAAGAGAG CCCAACCAG  
13601 CCAGGACTGT GGGCACAGTC TGGGCTGTTT GACTTCCCAT ATCTGAAAA  
13651 CCCAGAGAA AGCCAGCATA CTCTTGCTGG GGATGGCTGG GGAGAGGGCA  
13701 GTGGCAGAGA AAGGAGGGCA AGGGCAGGTG GTGAGATTCA ACATCCTTCC  
13751 AAAGCATTG CCAGAACCCG AAACCAATG GGACCCACC CCAGGAGAGC  
13801 GCCAGGGTGG AAGACAGAAG CTGTGTCTA CACACTGGGA GTATTACAGA  
13851 GAAGGGGTCT TGGCCAAGGC AGGGAGTACG CTGAATGTTG GGGGAATCCT  
13901 ATCTTCTCTT CTTGAGAACT CAGAACAAGG AAATGATGAC TTCAGGGCGA  
13951 CTCCCACCAC TTCTCCACC ACTTCTCTCC CCTGCCCTGT GGTCTGGGAG  
14001 CTATGTCAAG GACCTGCCTG TCATCCTCAT AGTTATAGGA GGCCACAGGC  
14051 CACCAGACAT GTGTCTCCAG TGCAAAAAAG CAGACACAG AGTCTGGGG  
14101 GTGAGACAG GACCCATCC TACCTTGGCT CTGCCCCGC CCCAGCAGGG  
14151 GCACCCCTCC AGGCCATGT GCCATTAGCA TTCTCTTATG TTTTCTCTT  
14201 CCTGCTTCAT CCAGTCGAAA TGGCTATGCA CTGGCCCAGG TCATGGACAC  
14251 TGACATTTAT GTGAATCCCA AGACCTATAA TAGGGTAGGT AATTCAAGCT  
14301 TATGACCTCC TTCTTTTGCT CTGCACCACC CCAAGAAGAG GTTGCTTTTT  
14351 AAAGCCAATA AAGACATTTT TGCAACTTGA GCTCAGTCTC CCTGTCACAG  
14401 GCCCAGGATA TCCAGGGGAT TAAATCATC ACGTACTGCT CCCCTCTCTA  
14451 CTTTGCCAAC TCAGAGATCT TCAGGCAAAA GGTCTCGCC AAGGTAAGGC  
14501 TCAGTCCCTG GCGACCAGAG GCTCTGGACA GAGAGTGGCC GGAAATGGA  
14551 AGCAGAAGGG CGGTGGGAGC TGAGAATAGG CCACTCCCAT AGAGGGTGGA  
14601 GGTCAAGATT GCTGTTGGCT CTCTCCCTGC AGACAGGCAT GGACCCCGAG  
14651 AAAGTATTAC TAGCCAAGCA AAAATACCTC AAGAAGCAGG AGAAGCGGAG  
14701 AATGAGGCCC ACACAACAGA GGAGGTCTCT ATTCATGAAA ACCAAGGTGA  
14751 ATGAAGGCCA GAAGCAGCCC CGTGCCCTGC TCTCCTGCCC ATTCTGATAC  
14801 TGCCCCCTGT TACTCATGGT ACCCTGGGGG CCCCCTTCC CACCCTGACA  
14851 GGCAAAGACA GAAAGTCTCT GGGAACACTG CCTGGTGGCC GCTGGGCATT  
14901 TTTCTTCTTT TTTTCTTTT TCTTTTATA GATGGAATTT TGCTCTGTC  
14951 ACCCAGGCTT GAGTGCAATG GCGTTATCTT GGCTCACTGC AACCTCCACC  
15001 TCTGGGGTTC AAGCGATTCT CCTGCCTTAG CCTCCCAAGT CGCTGAGATT  
15051 ACAGGTGCCA CCACACCAG CTAATTTTGT TATTTTATG AGATATTGGG  
15101 TTTACCATG TTGGCCAGGC TGGTGTCAA CTCCTGACCT CAGGTGATCC  
15151 ACCTACCTTA GCCTTCCAAA GTGCTGGGAT TACAAGCCTG AGCCACTGCG  
15201 CCCAGCCTGG GCATTTTCTT TCTTGATGA GGTGCTACCA TCTCCAGGG  
15251 AAGCCACTGA ACCCCCAAG CCCTTCTCCA TTTTCTGGCT AAGATAGGAC  
15301 ATGGCCCATG GACTTTTGAA CAACCCAGAG GGGGAACAGC AGTGAATTTT  
15351 CTGGGGAACC CAGGCAGCCC AGGGCTAGCA AGGCTGGGGT GGCCATGGCA  
15401 GTAATCCTTG TAATCCAGC ACTTTAGGAG GCCGAGATG GAGAATCACT  
15451 CTCATGAGTT CAGGAGTTCG AGACCAGCCT GCCCAACGTG GCGAAACGCT  
15501 GTCTCTACTA AAAATACACA AAAATTAGCC AGGCGTGGTG GTGGGCACCT  
15551 GTAATCCAG CTACTCAGGA GGCTGAGGCA CGAGAATCAC TTGAACCCGG  
15601 GAGGCAGAGG TTGCAGTGA CCGAGATAGT GCCACTGCAC TCCAGCCTAG  
15651 GCAACAGAGG GAGACTCTGT CTCAAGAAAT AAAGGAGCTC AGTGTCCCCG  
15701 GAGGGGCTTT CTCCAGAGA GAGTGGGCTT GAGGCTTCAG TGCCTCTCTT

FIGURE 3, page 5 of 20

```

15751 GGCTGGGTCC TCTGACTTTG TCTGGGTTGT AGGAGACCAA GTTTGCAGGC
15801 CCTGCCTAAG AAAGGGCTTT GGGAGAGGCC TCTCTGGTGG AGCTTTCAGG
15851 GTCTGTGTTC ACCATCACCG AGGCGAGTTA TTCCCCTACA CCTACACCCT
15901 CCATGCCCCT GCTTCAGTCA CAGCAAGGTC TGGCTCAGTC TGGTGGTCCC
15951 TGA CTCTGCC CACTGTCCCC ACCCTTCCAG ACTGTCTCCC TGCAGGAGCT
16001 GCAGCAGGAC TTTGAGAATG CGCCCCCAC CGACCCCAAC AACAACCAGA
16051 CCCC GGCTAA CGGCACCAGC GTGTCCTATA TCACCTTCAG CCCTGACAGC
16101 TCCTCACCTG CCCAGAGTGA GCCACCAGCC TCCGCTGAGG CCCCCGCGA
16151 GCCCAGTGAC ATGCTGGCCA GCGTCCCACC CTTCGTACC TTCCACACCC
16201 TCATCCTGGA CATGAGTGGA GTCAGCTTCG TGGACTTGAT GGGCATCAAG
16251 G CCTGGCCA AGGTGAGGCC CTCGGGGACA GCAAGCACCA CCCACTCCAC
16301 CCCCTCGCT CTGTCTCCA CATTCCTTT CCTGGGAGCC CTCATTTCAG
16351 GAAGCTGAGG GAGGAAGCTC ACTGGGGAGA CTAACAGCTC CTAGGAATCC
16401 CTCCTTTCCC CAGACGCCAC CAGGTTGAGA CATTCTCCAC AGAGCAGGCC
16451 CAGACGGCCC ATGACAATGA GTGGCGGGAC AAGTCTACCA GAGTTTCAGG
16501 CCCCTGTGCT CCCAACACCC CCAGCAGTGG CCATCCCAAG TCCCTCTCAG
16551 CCATCAGGAA CCCACCCAGG TTCTCTGAGG AGGGTCCAGT TTGGCTCCTG
16601 GTTCATGATC TGCTGCCCTT GTCCCTCAT CACCAGCCAC CCTAGGACAG
16651 GAGAAGAAAT AATACCACTG CCCCACACCA TCAGGCCAAA CAGAGAGCCC
16701 ACGGGACACC TTGAATGAAT GTATCCATCT GATAACTTTC CAGCAGCCAC
16751 CGCCAATGGC GGGAGTCAGC AAACCTCAGA GCTGGCTCAG ATAGAGGCAA
16801 GCCAGGGGAA CAATGGGCAC AGAGAGTGT CGGACTGCCT TCACCATCAA
16851 CCAGGCGCAG GGCAGGCCCC ATACCCAGCC TTGGGCCTCA GCCGGCTTCC
16901 TTAGCCAGGA TCTGGAGTCC AGGCCAGCCT TGGCTGAAGC TCTAGACTCC
16951 CTGAGCCTCC ATCTCCCTT GCAGCTTCTG TCTGAAGCCA CAAAGAAGTC
17001 TGAGAATCTA AGCTACTGAA AGAAAAGATC AGCCGGGCGT GGTGGCTCAC
17051 TCCTGTAATC CCAGCACTTT GGGAGGCCAA GGCAGGTGGA TCACAAGGTC
17101 AGGAGTTCAA GACCAGCTG GCCAACATGG TGAAACCCCG CCTCTACTAA
17151 AAATACAAA ATTAGCCAGG TGTGGTGACG GGCCCCTGTA GTCCCAGCTA
17201 CTCGGTAGGC TGAGGCAGAG AATTGCTTGA ACCCAGGAGG CGGAGGTTGC
17251 AGTAGCCCAA GATCGGCCA CTGCACTCCA GCCTGGGCAA CAGAGTGAAA
17301 CTCCATCTCA AAAGAAAAA AAAGAAAAA TCTAGCCCCA CAAGAAGGGG
17351 CCATGGTGAC TTTAAGTGCC CGCCACGTTG GCAAAAGTCC ATTTCCGCTC
17401 CACTTCCCAG AGAAACCGTC AGCCAACACT CCAGGGAGAA GTGGTGTGCT
17451 TTGCTGCTAT TTTTGTCTT GGCTGCTGG CTCTCAGGGT TGCTTATTTG
17501 TTTGGCTTCC CCTCTGAAGT ACGTTTTGTG AATCACTTTT GAGACCCACT
17551 CAGAACATTC CTTTCTTTT GCCTCCCTAC CCCAACAA CA TTCTAGCTG
17601 AGTCCACCT ATGGGAAGAT CGGCGTGAAG GTCTTCTTGG TGAACATCCA
17651 TGGTAAGAGA AAGAGGACAT TTAGGGACTG AAAGACTGGC AAGGAGTGTG
17701 GGGTAGGAAC AGGTGTTGGTGG GGTCTGAATA GTGAGGAGGT TGGAAACGAG
17751 AGCACCAGC TATCCCCAC AAGCTGCTGC CTGCTCATAA AAGCTTCAGG
17801 TACAAGTCCA AAGAGACTGG TCAGATTGCA TAAACATCCT AGGGGCCTTA
17851 GTGACAGAGT GGGGGTGAGG AGGTCATGGA GTTACAGAAG GACAGCTAGG
17901 ATTCTAATCT ACCCCATAAC TAATTTGCCA CGTATCCTTG GCCGAGTCAC
17951 TTTATCTCTC AAGGGATCTA TTTCTACCTA TGTAACCA GAGGGTTGAC
18001 TAGATGGATT TGGGGATCCT CTCCCAATCA GAAACTCTGT GAATCGATAT
18051 AGGCATAGAG CACACGGTAC CCTAATTCCC CAGGGAACAT ATAAATATGC
18101 AGTTTTGTAG GCATACAGCC TCCAAAGGGT GCATATACAC AGCCTCAAGG
18151 ACGTGGCCAC AGGGCAGCAG ACATTTACAT GACTAGCATG TACGCAAAGT
18201 GCAGAGATGT GGGAGCAAGT GCACACAGAC ACACAGGAGA ATGTGAAGGG
18251 GCACATACAC AACACCCAG CTCCCTGCAC TGGGTCAGAC CCCCTCAGC
18301 AGGGCTGCAG TTCCCAAGCT CCGCATGGCC ACGTTCGGGG AGAGAATCTG
18351 CAGTGGCAAT GACCTGCTAT GATATGTTCT GGAGTTAGAA GCAGTGGATT
18401 CTCCCCAACC TCACTGGACA CCCCCTTAGG AAACCATCTC TAGGATTAAG
18451 AGTAATCCAC ACAAATCTCC AATGCCACAC ATTGGAAGTT GCTGGAAGG
18501 TCTGGGAAAA CAAGAGGAAG GATGGGTCCT TGGGGGATAG AACTGGCAGC
18551 GGCCTCTTCA AGGATGGCTT AGGCTTTTCC ACTCGAATCA CCACAAAGTA
18601 CTGACTCCCT AAATCAAAT GCTTCTTCT GCTCTGGGTT GAAACTTCAG
18651 CATCCTCAAG TTCATGTTGC CCTCTGCCGT CCAGAACTGA TATTGCACTG
18701 CCAATGCCAT GGCCCTCAGA TACAGCAAGA GCTGGGACCT CAGGCCCTC
18751 CCATCCCTGC TCTGGTCTCA CTATCTTCCC CACCCCGAGC TCCAATCCAC
18801 AATGGCTGTT ATCTTTCTGA AGGTGATCTT TTCTCCTTCT AGCCAGGTG
18851 TACAATGACA TTAGCCATGG AGGCGTCTT GAGGATGGGA GTCTAGAATG

```

FIGURE 3, page 6 of 20

```

18901 CAAGCACGTC TTTCCAGCA TACATGACGC AGTCCTCTTT GCCCAGGCAA
18951 ATGCTAGAGA CGTGACCCCA GGACACAAC TCCAAGGGGT AAGGTTCTTG
19001 CACCTGGGGA ATCCTAGGCT CCAAGGCACT GAAATAGCAG GACCAAGAGG
19051 CATTATTAGA AAGAACACAG GAGAAGGTTT AAGTTCCAAT ATCAAGTCTG
19101 CCATTTTCAGT TTTCTGAATC TGTTTCCTTA TCTATAGAAT GAGCACCATC
19151 AACTAACATT ACCTACCTCT CTGCATTTTT CTTTTATTTT GTTTTAGGGT
19201 TAAATGATAA TTACATCTTT TGTGTCACTT GAAAGCACTT TGTGTATTGT
19251 AAAAATTCTT TATCAATATA AGTTTTCTGG TTGCACAAAC ACCCAAAGCA
19301 TAGTAGAGCA GGCCCACTCT GCTGGCATCG TTCCCTGCCT CCTCCTCATC
19351 TCTTTCTAAA GGGGGCTTTC GGAAGGGGAG GGGAGGGGAG TAAGCCTACC
19401 CATTTTAACT TACCGGAGCT TAGAGATTTC AGGCTGGTGA GGGATAAAGA
19451 GATTGGTGT GAGTTTTGTC TCAGCTTTT GACATTTAAT TTACTAGCTC
19501 AGTAAGTCAT ACAAATGGGA TACAAATAAC ACCATCTAAA ACTCCAGAAG
19551 ACTGGGGAGT CAGAAAAATC CTACCTCCTT GGGGTCCCTG CCCAGATCCC
19601 CAGTCATCTC TAGCCCTCAG GGTCCCCTCC CAGCTCAGCT CCTGCCCTTG
19651 GCCTCCCAG ACTCTTGTG TGCCCCAGCC CTGGGTAAAA ACCTCCCCTG
19701 CCCTCTGTGG GTCATAAGAA AGGCTTTTCT GGCCCTAGAG CAATGATTTG
19751 CTCTTTGCCT TCAGAGACTG ATGAAGGTGA AACCATCTGT TCTAAGTGCT
19801 GAAAGACTGC CCAGGAACAC ACAGGGCGCT GGCTCCTGCC CTCCATGCCT
19851 AGAGGGAAAC CCTGGGGAAA CAACGGGCTT TCCTGCTTCG TGAAATTTGT
19901 CCGCAGAGCA AAGAGGGAGA TTCTGGAGGA AGCTGCATTA GTTGTAGTG
19951 CCCTAATCAT GTTCAGCTAC TCTAGTTGGT ATGTATACTT GATTAGTCAT
20001 AGCACTTATA AATAATTTAT ATTTTATATA ATATATACTT ACATATTATA
20051 GACCAATTCAG AGATACAAAT CACACACATA AACACACACC TTTTCAACAG
20101 CATTGTGAGG GACAAAGCAG GCAAAGTGAG GCTGGTTATC AGACTTTAAC
20151 AGATTAGAAA ATATATTTCC AGGAGGACAG GAATTCCTCC AGGTCAGGCA
20201 GCTAGCCAAT AGTTTTTCTA AGCTGAGTAA AACCTTCCCT GCCTCTAACG
20251 GCCCACAAAG GAGGGAAGAC CGCGATACAC ACCTGTCTGG TATAAGGGGG
20301 AAGACCACAG CCGTGCTGTT TTTGTGAGGC AGGTAAGGGA AGGGGCAAGA
20351 GGATAAGTCA TGTGTCAGGA AGCAGCGTCC AACCAGAGCC GGCCACCTGT
20401 CCCTTTTCCT GCCACCATG ACCAACTTTG CTGTTCACTG ACTGAAGCTC
20451 ATTCTGCACT GGCTTCCTCC CTTCCAGGCT CCAGGGGATG CTGAGCTCTC
20501 CTTGTACGAC TCAGAGGAGG ACATTCGCAG CTACTGGGAC TTAGAGCAGG
20551 TGAGCTGAGG GAAGGGGCTG TGAGGGTGGG AGCAGGGCGA AGAGGGGAAG
20601 GATGGGGTCG CTGTCAAATA CAAGGCGTTC ACTCAGCTGT CTCACCTCCA
20651 GCCCAGAGCA GTCACATTCA AGGCCACAAA GATTTGTGGT CATCTTTGTT
20701 TTTTTTCTTT TCCTTTTCTT TTTTTTTTTT TTTTAATTTG AGACAAAGTC
20751 TCACTCTATC ACCAGACTG GAATGCAGTG GCATGATCTC AGCTCACTGC
20801 AACCTCTGCC TCCCGGGTTC CAGAGGTTCT CCTGCCTCAG CCTCCCGAGT
20851 AGCTGGGACT TCAGGCCTGC GCCCAGCTAA TTTTGTATT TTTAGTAGAG
20901 ACAGCTTTTC ACCATGTTGG CTGGGCTGGT CTCGAACTTC CGATCTCAAG
20951 CAATCTGCCT GCCTCGGTCT CCTAAGTGCC TGGATTACAG GCATAAGCCA
21001 CGATGCCTGG CCTTTGTTTT CATTCTTCTC ACTCCCTGAA AGGCATCGTG
21051 GGGAGAGGGT GAGTCACTGG ACCAAGTCCT AGAGAACCAG TATCTATTCT
21101 TATTCTCCAA CACATCACCC ACGTGACCCT GAGCAAGCCA CATACACCCT
21151 GGGCCCTAGT TTTTATCATC TGTGAAATTA GGGGAAACAT AGGTAATACC
21201 TGTCCCATCC ACCACACAAG ATTGGCAGGG CAGTCACTTG TTCTTTCATT
21251 AATTCAGCAG GTATTTATGG CGTACCTACT GTTTGCCTGA CACAGTTCAG
21301 GATGGGCACA TAGCAGTGAG CAAAACAAAG GCCTCTGCCT TTTAGAAACT
21351 TACGTTATGG TAGAATAGAT GGATTNNNN NNNNNNNNN NNNNNNNNN
21401 NNNNNNNNN NNNNNNNNN NNNNNNGTCT ACAAATGAAT TATTATTGCA
21451 TGTGGACAAG CCTTAAGAAC TAAAAAATAT GTGGCTGGGT GCAATGGTTC
21501 ACACCTGTAA TCCCAGCACT TTGGGAGGCT GAGGTGGGCG GACCACCTGA
21551 GGTCAGGAGT TTGAGACCAG CCTGGCCAAC ATGGCGAAAC CCCGTCTCTA
21601 CTAAGACAC AAAAATTAGC CAGGCGTAGT GGTGCATGCC TGAGTCCCA
21651 GCTACTCGGA AGTCTGAGGC ATGAGAATCA CTTGAACCTG GGAGGCAGAT
21701 GTTGCAGTGA GCCGAGATCG TGCCACTGCA CTCCAGCTTG GGTGACAGAG
21751 CTAGACTGTC TCAAAAACAA ACAAACCTAA AAGATATGTG
21801 GATATGAGGG ATCACCATCC CCATAGGGCC CCTGGATTAA CACCACCCCA
21851 CCAATGCCCT GAATTAAGAG AAACCAGATG ACTAGGTTTG GAGAAATCTG
21901 GCTTTGGGTC TATGAGAAAG AGTGTCTCTC TTTGTGCCTC TTCCATCTCT
21951 TTTTGACATT GAGCTCCATG GTGCTCTGAA TCCGTCTCTC ACAGTGCTGA
22001 TGGCAGGTGG GACAGATTAG AAAATAGAGC TGGAGCCACA GAGATTTGGC

```

```

22051 AGACTGATTT CGGTGCCCTC TTGGAATCTC CAGCACATTC CAAAAAGCCT
22101 GGATAGGACC AAAATAGCTT ATCAACGTGA GAAAGGACTT CAGAGCTTGT
22151 CTAAGTCCAA CCCTCATTTT ACCCAATGAG GAAAGTGAAG CTATTAGGGG
22201 GCGAGGGACA CGTGGGAAGT CACACAGCAC ACAGGAGGTG ATTCACATGT
22251 AGATTTTCAGC ACCTGCTCCT GCCACGCTGG ACTGGTTTAC CTCCTAGGCT
22301 GACCCTGCCT CTCCCCTGTT CACACACACT CTCGCACACA CACACACACA
22351 CACACACACA CACAGGTGCT TTGTTCTGGC CAGGGGTTC TAGGGTCACC
22401 TCTTGGTTGC AGCCACTGTG ACCCCAATG GTCTAACCTC TCTCTTCCCC
22451 TCCCACTTCC TTCCTGTGGT TCCTGCAGGA GATGTTCTGG AGCATGTTTC
22501 ACGCAGAGAC CCTGACCGCC CTGTGAGGGC TCAGCCAGTC CTCATGCTGC
22551 CTACAGAGTG CCTGGCACTT GGGACTTCCA TAAAGGATGA GCCTGGGGTC
22601 ACAGGGGGTG TCGGGCGGAG GAAAGTGCAT CCCCCAGAGC TTGGGTTTCT
22651 CTCTCCTCTC CCCCTCTCTC CTCCCTTCTC TCCCTCCCCG CATCTCCAGA
22701 GAGAGCCTCT CAGCAGCAGG GGGGTGCTAC CCTTACAGGA GTGAGAGTCT
22751 GGTGAGCCCA CTCTTCACCC GTCAGGCCCT GGCCGCAATG GACAAGCCTC
22801 CTGCTCACTC CACCCACCCC ACCTCTGCCC TGTCCTTGGC AGCTGAAGGA
22851 CACCTTGACT TCCAGCTTTT ACGAGTGAGC CAAAAACAGA AGGACAAGTA
22901 CAAGTGTGCT GGCCTGTGCT ACAAGCTTCA AAAAGTGTCC CAGAGCCCAC
22951 ACGGCTCGGT GTCAGATGCT GTCAGGCTGT CACGGACATA GGGATAAACT
23001 TGGTTAGGAC TCTGGCTTGC CTTCCCCAGC TGCCTCAACT CTGTCTCTGG
23051 CAGCTCTGCA CCCAGGGACC ATGTGCTCTC CACACCCAGG AGTCTAGGCC
23101 TTGGTAACCTA TGCGCCCCCC GTCCATCATC CCCAAGGCTG CCAAACCAC
23151 CACTGCTGTC AGCAAGCACA TCAGACTCTA GCCTGGACAG TGGCCAGGAC
23201 CGTCGAGACC ACCAGAGCTA CCTCCCCGGG GACAGCCCAC TAAGGTTCTG
23251 CCTCAGCCTC CTGAAACATC ACTGCCCTCA GAGGCTGCTC CCTTCCCTG
23301 GAGGCTGGCT AGAAACCCCA AAGAGGGGGA TGGGTAGCTG GCAGAATCAT
23351 CTGGCATCCT AGTAATAGAT ACCAGTTATT CTGCACAAAA CTTTTGGGAA
23401 TTCCTCTTTG CACCCAGAGA CTCAGAGGGG AAGAGGGTGC TAGTACCAAC
23451 ACAGGGAAAA CGGATGGGAC CTGGGCCCAG ACAGTCCCCC TTGACCCAG
23501 GGCCCATCAG GGAAATGCCT CCCTTTGGTA AATCTGCCTT ATCCTTCTTT
23551 ACCTGGCAAA GAGCCAATCA TGTTAACTCT TCCTTATCAG CCTGTGGCCC
23601 AGAGACACAA TGGGGTCCTT CTGTAGGCAA AGGTGGAAGT CCTCCAGGGA
23651 TCCGCTACAT CCCCTAATG CATGCAGATG TGGAAGGGG CTGATCCAGA
23701 TTGGGTCTTC CTGCACAGGA AGACTCTTA ACACCCTTAG GACCTCAGGC
23751 CATCTTCTCC TATGAAGATG AAAATAGGGG TTAAGTTTTC CATATGTACA
23801 AGGAGGTATT GAGAGGAACC CTACTGTTGA CTTGAAAAATA AATAGGTTCC
23851 ATGTGTAAGT GTTTTGTAAG ATTTCAAGTG AAATGCACAG AAAATCTTCT
23901 GGCCTCTCAT CACTGCTTTT CTCAAGCTTA TTCAGCTTAA CAACCCCTTC
23951 CCTAACAGGT TGGGCTGGCC CAGCCTAGGA AAACATCCCC ATTTCTAACT
24001 TCAGCCAGAC CTGCGTTGTG TGTCTGTGTG TTGAGTGAGC TGGTCAGCTA
24051 ACAAGTCTTC TTAGAGTTAA AGGAGGGGGT GCTGGCCAAG AGCCAACACA
24101 TTCTTGGCCC AGGAGCATTG CTTTTCTGTG AATTCATTAT GCCATCTGGC
24151 TGCCAATGGA ACTCAAACT TGGAAGGCGA AGGACAATGT TATCTGGGAT
24201 TCACCGTGCA CAGCACCCGA AGTGCCAAAT TCCAGGAGGA CAAGAGCCTT
24251 AGCCAATGAC AACTCACTCT CCCCTACTCC ACCTCCTTCC AAGTCCAGCT
24301 CAGGCCAGG AGGTGGGAGA AGGTCACAGA GCCTCAGGAA TTTCCAAGTC
24351 AGAGTCCCCT TTGAACCAAG TATCTAGATC CCCTGAGGAC TTGATGAAGT
24401 GATCCTTAAC CCCCAGGTAA TCATTAAACC CCAGACCAGC CTCAGAACTG
24451 AAGGAGATTG TTGACCCAGT GACCTGGAGT TGAGGCTCAG GGAGAGATCT
24501 GCCACATGTC TGAGGGTTGC AGAGCC

```

**FEATURES:**

```

Start: 1997
Exon: 1997-2121
Intron: 2122-4732
Exon: 4733-4872
Intron: 4873-5004
Exon: 5005-5115
Intron: 5116-5781
Exon: 5782-5957
Intron: 5958-7770
Exon: 7771-7935
Intron: 7936-8470

```

Exon: 8471-8623  
 Intron: 8624-8917  
 Exon: 8918-9000  
 Intron: 9001-9777  
 Exon: 9778-9925  
 Intron: 9926-10221  
 Exon: 10222-10335  
 Intron: 10336-10539  
 Exon: 10540-10617  
 Intron: 10618-11197  
 Exon: 11198-11293  
 Intron: 11294-13338  
 Exon: 13339-13445  
 Intron: 13446-14214  
 Exon: 14215-14284  
 Intron: 14285-14400  
 Exon: 14401-14493  
 Intron: 14494-15980  
 Exon: 15981-16262  
 Intron: 16263-17597  
 Exon: 17598-17652  
 Intron: 17653-18842  
 Exon: 18843-18988  
 Intron: 18989-20477  
 Exon: 20478-20549  
 Intron: 20550-22478  
 Exon: 22479-22523  
 Stop: 22524

**CHROMOSOME MAP POSITION:**

Chromosome 1

**ALLELIC VARIANTS (SNPs):**

DNA				Protein		
Position	Major	Minor	Domain	Position	Major	Minor
48	C	G	Beyond ORF (5')			
132	G	A	Beyond ORF (5')			
724	A	C	Beyond ORF (5')			
1558	C	G	Beyond ORF (5')			
1577	A	G	Beyond ORF (5')			
2487	C	A	Intron			
2634	T	C	Intron			
4352	A	G	Intron			
5157	A	C	Intron			
5658	A	T	Intron			
5945	T	C	Exon	180	T	T
6281	C	T	Intron			
6452	G	C	Intron			
6610	T	G	Intron			
7247	T	C	Intron			
7360	A	G	Intron			
7644	A	T	Intron			
8127	A	C	Intron			
8317	G	A	Intron			
9079	G	A	Intron			
9537	G	T	Intron			
12302	C	G	Intron			
12354	C	T	Intron			
12487	C	T	Intron			
13198	-	A	Intron			
13257	A	G	Intron			
14541	G	A	Intron			

002227" 69564260

14545	A	G	Intron			
15041	C	A	Intron			
15053	A	C	Intron			
15065	A	G	Intron			
15108	A	C	Intron			
16274	-	G	Intron			
17424	C	T	Intron			
17627	G	A	Exon	657	V	V
18427	T	C	Intron			
18813	C	G	Intron			
19035	T	C	Intron			
19182	T	C	Intron			
19508	-	G C	Intron			
19571	T	G C	Intron			
20147	T	G	Intron			
20180	G	A	Intron			
20584	A	T	Intron			
20717	T	C	Intron			
20894	A	G	Intron			
21787	-	A C	Intron			
22264	T	C	Intron			
22338	-	C A	Intron			
23363	T	C	Beyond ORF (3')			
23688	G	A	Beyond ORF (3')			
24210	A	C	Beyond ORF (3')			

Context:

DNA

Position

48	CTGGGTTCTATGTGGGGAGGTCATGCTCCCCACTCATTGAGCCCCC [C, G] CAGGCAAACACCTGGACAGCCAGACCCATGCAGACTCTGGAGCAGGTGGAGAGGAAGAG TGAGACCACCCCGCCTCACGGGCGGTGAAGGGCCGGCAGCCTCTGAATAGTCTCTGCTAG GAGGTAGAAAGCACCTCCCATCTTAATCATAGTAATCATCGCCACTACCATTTACTGGG TGCTATAAAAGGCCAGCCTCTTCATACACATGATCTCACTGAATCCTCATAGCATCTGC CTGCGACTGTTATTATCCCCATTTACAGATGAAGAACTGAATCTTTGAACCCAGGTCAT
132	CTGGGTTCTATGTGGGGAGGTCATGCTCCCCACTCATTGAGCCCCCCCAGGCAAACCA CTGGACAGCCAGACCCATGCAGACTCTGGAGCAGGTGGAGAGGAAGAGTGAGACCACCC GCCTCACGGGC [G, A] GTGAAGGGCCGGCAGCCTCTGAATAGTCTCTGCTAGGAGGTAGAAAGCACCTCCCATCT TAATCATAGTAATCATCGCCACTACCATTTACTGGGTGCCTATAAAAGGCCAGCCTCTTC ATACACATGATCTCACTGAATCCTCATAGCATCTGCCTGCGACTGTTATTATCCCCATTT ACAGATGAAGAACTGAATCTTTGAACCCAGGTCATCTGGCTCTCAAACCTGTGTCTGTTT TCCCTAAGCCACCCGGTCTCTCATTTCTCCCACTGAAATGTCTCACATGCCATTGCCCTT
724	ATTGCCCTTACTCATTTCTGCCCATGTCTCCTCCAAAACACCATTTATCAATTCGCTCAA CAAGTATGTGTTGAGTACACACTAAGGGCCAGGCGAGGGGCTGGGCACAGGCGCTGGGGG TAGGTTCAATTCCTCCACCTTCGCTTCTGCTGGGTATCACCTGTGGGGTCTTGCCGGGCAT CCCACCCTCACCTGTAGTTCAAGTGGACCTTGGGATCCCAAGACCAAATGAATGGAATGC ACCAGCCAGCCTTCACCAACTTGAGCACAATCTTATTATAATAGAACTCACATTTGC [A, C] TCACACTTTACATTTTACACAACCCCTTCTTATCCATTAACCTCATTTGATCTTCACAACA ACCTGTGAGATATGTCTGTTACTCCCACTTTAGTGATACAGAATCTGAGGTTTGAAAAG TAATGCTGACCATTCTGCCTCATTAATAAAAGCAGGATTAACCCAGGCTCCTGGACCCTT CCACAAAAGGCATTAAGCAACCTGCTCCCTCTGACAACCTCCCTGTACCCAGGCTCT CCTCTGGGAAGTTGGGGGCATCTCTAGCCCCAAGTAGTTACTCATTTTCAACCCCATCT
1558	TCAGCTCTGCCCATCTCAGCTCCTGGAACGTGAGCCAGGTTGCGCAAAAAGTGAGGAGGA GAGGAGCGGCAGTACACAAGGGTGGGGGAAAGATTAGGCACAGGAAGCCGTGGGAGAGAG AGCCGGCAGGTGGACCATCCTGGTTTCCCCACACACACCATTTGTCCTTGGGAAACCTG

FIGURE 3, page 10 of 20





CAAGGATGGGAGGTGTGGCAAAGGGGCTCGGGAGATTTTCCATCTGCATTCTCCTGGAG  
 TTGTTCTCGGTGAGTCTAGGGGAATGGTCACTGTGAATGTCATTTCCAGGTCTCGGTG  
 ACCTTGGAGAAACCACTGAGCCTCTTTGAGTTCAGTTAGCATTACCTGTTCCATCTTCT  
 CCTAGGAATGAGAGGAAGACTTAGCAGAACAAGATATACCATATGCTATAACATGCTTAA  
 ACAGATGTGAGAAATCACCATCTAACTCCCTGGTTGGTCCCAGCCGGCCACTACAGGGAC

5658 TTAGCAGAACAAGATATACCATATGCTATAACATGCTTAAACAGATGTGAGAAATCACCA  
 TCTAACTCCCTGGTTGGTCCCAGCCGGCCACTACAGGGACATTTGACTTCTCTGGTGCT  
 AAGTGAGATGGAGGAAAGCCTGGTCACAAGGGCTGGTTTCTGGTTCAGGCTCTGCTTATA  
 TTTCTTATTTCTGAGTTCATTTTCTCACGTGTCTGTATGACAATATTGACCATTGGGGT  
 AAAAGCACCTTGAAGCATAGATCATGGTTAGAGTGAGTGGTTGTTATTATTGTGTTGG  
 [A, T]  
 GAAGAGCCTTGGAGGTGCAGGGATCCATCCCCCTGGGGTGGGAAGCATTCCTGGGCCCC  
 TTTCTGGTTTCCATCGGTGTGGTTCAAACCTCTGATTTTGGTGGCTGGGTGGGGACCA  
 CAGGTACCTTTGCCGTTATCAGCATCCTGGTGGGTAAACATCTGTCTGCAGCTGGCCCCAG  
 AGTCGAAATTCAGGTCTTCAACAATGCCACCAATGAGAGCTATGTGGACACAGCAGCCA  
 TGGAGGCTGAGAGGCTGCACGTGTGAGCTACGCTAGCCTGCCTCACTGCCATCATCCAGG

5945 ATTATTGTGTTGGAGAAGAGCCTTGGAGGTGCAGGGATCCATCCCCCTGGGGTGGGAAG  
 CATTCCTGGGCCCCCTTTCTGGTTTCCATCGGTGTGGTTCAAACCTCTGATTTTGGTGGC  
 TGGGTGGGGCACCACAGGTACCTTTGCCGTTATCAGCATCCTGGTGGGTAAACATCTGTCT  
 GCAGCTGGCCCCAGAGTCGAAATTCAGGTCTTCAACAATGCCACCAATGAGAGCTATGT  
 GGACACAGCAGCCATGGAGGCTGAGAGGCTGCACGTGTGAGCTACGCTAGCCTGCCTCAC  
 [T, C]  
 GCCATCATCCAGGTGAGGGGGCAGCCCCCAACCCTGCTAGAAGGGCATCAGACCACCCTG  
 CCCCTCCCTCAAAGCCTTAGCTTTGATGCTAAATCTGATTTAGGGGGCTGGGTGTGGAGG  
 CTCATGCCTGTAATCCAGCACTTTGGGAGGCTGAGGAGGTTGATCACTTGAGGTGAGG  
 AGTTTGAGACCACCTTGACCAACGTGATGAAACCCATCTCTACCAAAAATACAAAATA  
 ATCCAGGCTTGGTAGTATGCGCCTGTAGTCCCACCTACTCAGGAGGCTGAGGCAGGAGAA

6281 GCTAGAAGGGCATCAGACCACCCTGCCCTCCCTCAAAGCCTTAGCTTTGATGCTAAATC  
 TGATTTAGGGGGCTGGGTGTGGAGGCTCATGCCTGTAATCCAGCACTTTGGGAGGCTGA  
 GGAGGGTGGATCACTTGAGGTGAGGAGTTTGAAGCCACTTGACCAACGTGATGAAACCC  
 CATCTCTACCAAAAATACAAAATAATCCAGGCTTGGTAGTATGCGCCTGTAGTCCCACC  
 TACTCAGGAGGCTGAGGCAGGAGAATCACTTGAATCCGGGAGGCAGAGGTTGCAGTGAGC  
 [C, T]  
 GAGATCGCGCCACTGCACTCCAGCCTGGGTGACAGAGCGAGACTCCGTCTCAAAAAAAA  
 AAAAAAAAAAAAAAAAAACCAAGTTAGGGCTCACCTCCTCCTCCCATCCAGG  
 GCTAAAGTGAACCTTGAAAATTAACAGTATCTCCTCATCTGCATGTAGCAGGACCATA  
 AAAAAACAACAGCTGTACCTGGTTAACTGTCTGAGCTTTAACTGTAAAAGACTCAC  
 AGCCTCTCTCCATTATCCCGTGGAGAAACCAACTCTCTGCCAGCATAGTCTTGCACT

6452 ATGAAACCCCATCTCTACCAAAAATACAAAATAATCCAGGCTTGGTAGTATGCGCCTGT  
 AGTCCCACCTACTCAGGAGGTGAGGCAGGAGAATCACTTGAATCCGGGAGGCAGAGGTT  
 GCAGTGAGCTGAGATCGCGCCACTGCACTCCAGCCTGGGTGACAGAGCGAGACTCCGTCT  
 CAAAAAAAAAAAAAAAAAAAAAAAAAACCAGTTAGGGCTCACCTCCTCCTCCTCC  
 CCATCCAGGGCTAAAGTGAACCTTGAAAATTAACAGTATCTCCTCATCTGCATGTAGCA  
 [G, C]  
 GACCATACAAAAACAACAGCTGTACCTGGTTAACTGTCTGAGCTTTAACTGTAA  
 AAGACTCACAGCCTCTCTCCATTATCCCGTGGAGAAACCAACTCTCTGCCAGCATAGTC  
 TTGCAGACTGCTAATTTTCTTAACATCCCTCACTCCGCTCCAGCCTCCTCTGCTCCAAG  
 CCACAGCAGCAGTTGCACAACATAAATTGAGCTTCTGCAAAATGGTTGCAAGGATTCTGC  
 TAGGTTTTATGAAGGAAGCACAAATGACAGAATGCAAGAGCAAAACACAGTCCCAGAG

6610 GTGACAGAGCGAGACTCCGTCTCAAAAAAAAAAAAAAAAAAAAAAAAAAACCAGTTA  
 GGGCTCACCTCCTCCTCCTCCCCATCCCAGGGCTAAAGTGAACCTTGAAAATTAACAGT  
 ATCTCCTCATCTGCATGTAGCAGGACCATAAAAAAAAAACAACAGCTGTACCTGGTTAAAC  
 TGCTCTGAGCTTTAAACCTGTAAAAGACTCACAGCCTCTCTCCATTATCCCGTGGAGAAA  
 CCAACTCTCTGCCAGCATAGTCTTGACAGACTGCTAATTTTCTTAACATCCCTCACTCC  
 [T, G]  
 CTCCAGCCTCCTCTGCTCCAAGCCACAGCAGTTCACACAACATAAATTGAGCTTCTGC  
 AAATGGTTGCAAGGATTCTGCTAGGTTTTATGAAGGAAGCACAAATGACAGAATGCA  
 AGAGCAAAACACAGTCCCAGAGAGCGCCTTTTCATTCATTCATTTCGTTTTGTGCC

FIGURE 3, page 12 of 20

AAGAACTAGGCTAAACCTGGGATACAAAGATAAGTAAGAAAGAGGTCCAATTACACAAGT  
TGCTCACAGCCCAGCAGAGGAAGGAGCCATGTCAACAGATAAATTTGTATGCAGTGAGAT

7247 GACACAGAGCAGAGTCACGGAGGACCTCAAAGAGGAGGTGACACTCCACCTCTCTTAAAG  
GATGAGAACTTAACCAGGAACAAGGTATACAGAGGATGGTCCAGGCAGAAGGGAACAGTG  
CCTAAAAACACTGAGGCCTGAGAGAGTGTGATCTGCGCAGGCAAAGTAAGGGGCTTGGTG  
TGGCTGGAGGGTAGAGGGCCCAGAAGAGGATGGAAAAGTAGGCAGGAGCCAGACAATGAG  
ATCTGGGGTCTGTTCTCTGACAGCGACTTTGGGTCTGATTGGCAGTTTATAAGGATCGTT  
[T, C]  
GGGCTACACAATGATGAGTGGGAGGTGGATTAGAATCAAGGCAGGGGACCTGTTGGGAGA  
CTCTGCAGAGGCCAGGCAGGAATAATGCAGGCGAAGACCAGGTAGAGAAAAGAGATGGGG  
CTGGACTTGAAAAGAATGTTTTACCAGGAGCTTGGTGATAGACTGGATGTGGGAGGTAAG  
GGAGGATGACTCTCAAGTTTTTGGTTGGGCAACCAGGTTAATGATGGTGTCAATTTACTGA  
GAGAGAAAACACTGGGGGAGGACTAGACTTATTTTACAGATAAGCCAAAGCCAGAGAGGT

7360 AACAGTGCCTAAAAACACTGAGGCCTGAGAGAGTGTGATCTGCGCAGGCAAAGTAAGGGG  
CTTGGTGTGGCTGGAGGGTAGAGGGCCCAGAAGAGGATGGAAAAGTAGGCAGGAGCCAGA  
CAATGAGATCTGGGGTCTGTTCTCTGACAGCGACTTTGGGTCTGATTGGCAGTTTATAAG  
GATCGTTTGGGCTACACAATGATGAGTGGGAGGTGGATTAGAATCAAGGCAGGGGACCTG  
TTGGGAGACTCTGCAGAGGCCAGGCAGGAATAATGCAGGCGAAGACCAGGTAGAGAAAG  
[A, G]  
GATGGGGCTGGACTTGAAAAGAATGTTTTACCAGGAGCTTGGTGATAGACTGGATGTGGG  
AGGTAAGGGAGGATGACTCTCAAGTTTTTGGTTGGGCAACCAGGTTAATGATGGTGTCA  
TTACTGAGAGAGAAAACACTGGGGGAGGACTAGACTTATTTTACAGATAAGCCAAAGCCA  
GAGAGGTGATGTGACAGAAAGGCCCATGCTCTAAAGGAGCTGAAGGTCTGATGGCAGCCA  
TGTAAGGCACAGTGAAGGGCAGGTGAAGGTCACAGATGGTCCAATTCCTCAAGCTACTG

7644 GACCAGGTAGAGAAAAGAGATGGGGCTGGACTTGAAAAGAATGTTTTACCAGGAGCTTGGT  
GATAGACTGGATGTGGGAGGTAAGGGAGGATGACTCTCAAGTTTTTGGTTGGGCAACCAG  
GTTAATGATGGTGTCAATTTACTGAGAGAGAAAACACTGGGGGAGGACTAGACTTATTTTA  
CAGATAAGCCAAAGCCAGAGAGGTGATGTGACAGAAAGGCCCATGCTCTAAAGGAGCTGA  
AGGTCTGATGGCAGCCATGTAGAGCACAGTGAAGGGCAGGTGAAGGTCAAGATGCTCCA  
[A, T]  
TTCCCTCAAGCTACTGCTACGCTAGGACTGCACGGAGCTCCAGACCTGCGTGTGTGTGGG  
GCGGGTCGTTGGAAGTGTGAACCAATTGGTCTTCCGCCACCAACCACCTTTTCCTCC  
TCTCAGATGGGTCTGGGCTTCATGCAAGTTTGGCTTGTGGCCATCTACCTCTCCGAGTCC  
TTTCATCCGGGGCTTCATGACGGCCGCGGCTGCAGATCCTGATTCGGTGTCTCAAGTAC  
ATCTTCGGACTGACCATCCCCTCTACACAGGCCAGGGTCCATCGTCTTTGTGAGTCTG

8127 CATCCGGGGCTTCATGACGGCCGCGGCTGCAGATCCTGATTCGGTGTCTCAAGTACAT  
CTTCGGAAGTACCATCCCCTCCTACAGAGGCCAGGGTCCATCGTCTTTGTGAGTCTGGG  
GATGCACCCCTGCCATTGGAGCAAGGCTCCAGCAGACACATGAGGAGGATGTACTGTTTT  
AAGATGTGCTGAGCTCCTCATTGCAAGGGCTGGCTTAGCTGTTGTTAGAGAGGATTCTG  
AGGGGGTTTCTGTCTTGGGAGGTCAAAGTCATGACTCACAGAGGTTCTTGGTAGTTAAT  
[A, C]  
CCTGCAGAAAAGAGCTGTACATTCTCCGCCAGTTCCTCCATTCTAGTGCCTCAACCCCTCC  
CTGCCTGGAAAGTCTGCCTTATGTCTAATCTCCATCCCTCCTCCTCAGCCCAAACCTCT  
TCTAAAGAAAAAGAAAGCATTCCCTTTCTAGCACAAAGTTCCTCATGTGCCTTTTGGGAAA  
GGGCGGTGGGCGACGGGACAGGGTTCCTGATCAGGGTTTTAATTCTGTCTTGGTGTGCCT  
CCATTAGCTTTGATGGCATCCCTTCCCTGGGTGAGACACCCAAAGGTGGGGTATTATGGG

8317 GAGCTCCTCATTGCAAGGGCTGGCTTAGCTGTTGTTGAGAGAGGATTCTGAGGGGGTTTC  
TGTCTTGGGAGGGTCAAAGTCATGACTCACAGAGGTTCTTGGTAGTTAATACCTGCAGAA  
AAGAGCTGTACATTCTCCGCCAGTTCCTCATCTAGTGCCTCAACCCCTCCCTGCCTGGA  
AAGTCTGCCTTATGTCTAATCTCCATCCCTCCTCCTCAGCCCAAACCTCTTCTAAAGAA  
AAAGAAAGCATTCCCTTTCTAGCACAAAGTTCCTCATGTGCCTTTTGGGAAAGGGCGGTG  
[G, A]  
CGACGGGACAGGGTTCCTGATCAGGGTTTTAATTCTGTCTTGGTGTGCCTCCATTAGCTT  
TGATGGCATCCCTTCCCTGGGTGAGACACCCAAAGGTGGGGTATTATGGGAAGAAGGGGT  
GGGAGCCTGTGAGCATGATGCTCTTCCCCAGACCTTCATTGACATTTGCAAAAACCTC  
CCCCACCAACATCGCCTCGCTCATCTTCGCTCTCATCAGCGGTGCCTTCCCTGGTGTCTG  
GTGAAGGAGCTCAATGCTCGCTACATGCACAAGATTTCGCTTCCCATCCCTACAGAGATG

9079 TTCTACTGCTCTAATAATTCCCCCTAAGGAGGCAGGGGAGTGGGATTAGGGTCCCCAGA  
GAAAAGGGAGACTTGAGAGAGACGCCTGCCCTGGCCCCACCTTAGGGCCAATCCCCATTC  
TCCACTCTGGGGTTTGCAGGTGGTGGTGGCAACAGCTATCTCCGGGGGCTGTAAGATGCC  
CAAAAAGTATCACATGCAGATCGTGGGAGAAATCCAACGCGGGTGAGTCCAGGTGGCCCA  
GAAGCCTGGCCACCCGCACCTCATGCCCCACTAAGGCCTGAGCTCGGAGAGGGAGACAA  
[G, A]  
ATGAACTCTATGAAAGTGCAGTCGAAACTGTATGACACTGACCATGTATGAATTATTACT  
ATTACCGTTTCTGAGAAGGGCCGCACAACCAGCCAATGTAGGCTATTTATGAGAAATG  
AGTCTTAACTGCCACACTCCCCTTATAAATCTCATTCAACTGATGCTGTTAAACAAAGCC  
TCTCTGAACAGCCGCTTGTCTGGCTCTTGCCTTGCTCTAATGCATTGGTCTTTGTCCAT  
GTAGAAAGGGAACATTAGGTTCAACCAGATTCTGAAGCATCCACTCTGTGCCAGGCAC

9537 AACTGATGCTGTTAAACAAAGCCTCTCTGAACAGCCGCTTGTGGCTCTTGCCTTGCTC  
TAATGCATTGGTTCTTGTCCATGTAGAAAGGGAACATTAGGTTCAACCAGATTCTATGA  
AGCATCCACTCTGTGCCAGGCACCATGCTGGGCCCTGGGAGGAGAGGGGTGACGCTTGTC  
CTGCAGGGTTGGAACAGGCAAGGGAGGGAAGACCACATAGCACCAAAGGTCTAGGGGTCT  
GTGGACTCGTGAGCATACAGGGTTGAGAATCTGGGAGTTAAACAAACGAGGCCCTACCACA  
[G, T]  
ACTGGCCCGGGACCTTGGGCAAGTTAGGTTCTCTCAGCCTCAGTTTCTCCTTTGTAA  
ACAGGAGTGATGGTCCCTACCCTATGGGGTGGTGGTGGAGGATTGAGTGGATGGGATAA  
CTTAGGCAAAGATCCCGGCACACCATGGGGGCTGGCTGGTCCCTGTGGGCTGGTGAAGG  
ACTTGCTGCCCCCTCCCACTCACACCTTGGGTTCTGCCTCCTTGGCTCCTCGGCAG  
GTTCCCCACCCCGGTGTCGCCTGTGGTCTCACAGTGAAGGACATGATAGGCACAGCCTT

12302 AGCCCCACCATAACCTATGGGAGAGGATTTACTAACTTTCTTAACGTGATGAAACCAA  
GGCTCAGAATGGTTAAGTAAATTGTCAAAGGCCACAGAGGTAGGGAGTGGTAGAGTCTGG  
ATTAAACTCCAAGTCTGGAAGTCCAGACCTCTAGGCTGTACTGTCTCATAGGGAAGGCA  
GTCTCACCACCTAGGGCAGAGAAGAAAATCCTTAAAGCCAGAGAAGTGAGTGGCTCATC  
TGTGCTACCCAGAGAGACAGTGATGAGGACAGGGAGAAAAATTATACCTCAGTTCCAG  
[C, G]  
CCAAGGATCTGCTTTGACCATAACCCAACAAGCCCCGCTATGGTGGTATTTCTTAGGT  
TCATATGGCGGCTTTTGTTCATTTGATCTTCACAGCAATTCTCTACAGGAATCTGGGC  
AGATTTATTTCTTTAGAGGAATTTCCAGGTCTTAAATCTATAGGGGGCAACTATCAAA  
ACTTCACCAATGTGCCCCCTACCCACACACAAAACAGGCCCCAGCCGATCAGAAAG  
CACTGCTGAGCTCCTGTGAGGGCCACGCAGCTCGCTGTGAGACAGAGAGAGGGAACCTCA

12354 GAAACCAAGGCTCAGAATGGTTAAGTAAATTGTCAAAGGCCACAGAGGTAGGGAGTGGTA  
GAGTCTGGATTAAAACCTCCAAGTCTGGACTCCAGACCTCTAGGCTGTACTGTCTCATAG  
GGAAGGCAGTCTCACCACCTAGGGCAGAGAAGAAAATCCTTAAAGCCAGAGAAGTGAGT  
GGCTCATCTGTGGTCAACCCAGAGAGACAGTGATGAGGACAGGGAGAAAAATTATACCTCA  
GTTCCAGCCCAAGGATCTGCTTTGACCATAACCCAACAAGCCCCGCTATGGTGGTATT  
[C, T]  
CCTTAGGTTTCATATGGCGGCTTTTGTTCATTTGATCTTCACAGCAATTCTCTACAGGA  
ATCTGGGCAGATTTATTTCTTTAGAGGAATTTCCAGGTCTTAAATCTATAGGGGGCAA  
CTATCAAAACTTCACCCAATGTTGCCCTACCCACACACAAAACAGGCCCCAGCCGCA  
TCAGAAAGCACTGCTGAGCTCCTGTGAGGGCCACGCAGCTCGCTGTGAGACAGAGAGAG  
GGAACCTACATTTATTGATCACCTACTGAGCATCCATCACTAGGCTAGGACCGTCACATT

12487 ACCCACCTAGGGCAGAGAAGAAAATCCTTAAAGCCAGAGAAGTGAGTGGCTCATCTGTGG  
TCACCACAGAGACAGTGATGAGGACAGGGAGAAAAATTATACCTCAGTTCACAGCCAA  
GGATCTGCTTTGACCATAACCCAACAAGCCCCGCTATGGTGGTATTTCTTAGGTTTCAT  
ATGGCGGCTTTTGTTCATTTGATCTTCACAGCAATTCTCTACAGGAATCTGGGCAGAT  
TTATTTCTTTAGAGGAATTTCCAGGTCTTAAATCTATAGGGGGCAACTATCAAACTT  
[C, T]  
ACCCAATGTTGCCCCCTACCCACACACAAAACAGGCCCCAGCCGATCAGAAAGCACTG  
CTGAGCTCCTGTGAGGGCCACGCAGCTCGCTGTGAGACAGAGAGAGGGAACCTCACATT  
ATTGATCACCTACTGAGCATCCATCACTAGGCTAGGACCGTCACATTCTTAACTTTGA  
ATCCTTTCATGAGGTAGGCATTATTATTTCTCTTTTGTTCACATAGCCATTAAAGAACA  
AAATTTGGGGCTGGGTGTGCTGACTCACACCTGTGATCTAGCACTTTAGGGGGCTGAGGC

13198 CTAATATTAGGAAGGTTAGGCGGGAGCACAACCTTGGGTTCCAGGGTTTGGAGCTCCAG  
TGAGCTGATCTTGCCACTGCACTACAGCCTGAGCAACAGAGCAAGACCCTGTGACTCCAA  
AAACAAACAAACACACATTTGAACCCAAACAGATCTGACCCAAGATGCATGCTCTTA

09749589-123800



GGTGCCACCACACCAGCTAATTTTGTATTTTGTAGATATTGGGTTTCACCATGTTG  
GCCAGGCTGGTGTCAAACCTCCTGACCTCAGGTGATCCACCTACCTTAGCCTTCCAAAGTG  
CTGGGATTACAAGCCTGAGCCACTGCGCCAGCCTGGGCATTTTCTTCTGGATGAGGT  
GCTACCATCTCCCAGGGAAGCCACTGAACCCCCAAGGCCCTTCTCCATTTTCTGGCTAAG  
ATAGGACATGGCCATGGACTTTTGAACAACCCAGAGGGGGAACAGCAGTGAATTTCTCTG

15065 CAGCCCCGTGCCCTGCTCTCTGCCCATTCTGATACTGCCCCCTGTTACTCATGGTACCC  
TGGGGGCCCCGCTTCCCACCTGACAGGCAAAGACAGAAAGTCTCTGGGAACACTGCCTG  
GTGGCCGCTGGGCATTTTCTTCTTTTTTCTTTTTCTTTTTTAGAGATGGAATTTTGCT  
CTTGTCACCCAGGCTTGAGTGCAATGGCGTTATCTTGGCTCACTGCAACCTCCACCTCTG  
GGTTCAAGCGATTCTCCTGCCTTAGCCTCCCAAGTCGCTGAGATTACAGGTGCCACCAC  
[A, G]  
CCCAGCTAATTTTGTATTTTGTAGATATTGGGTTTCACCATGTTGGCCAGGCTGGTG  
TCAAACCTCCTGACCTCAGGTGATCCACCTACCTTAGCCTTCCAAAGTGCTGGGATTACAA  
GCCTGAGCCACTGCGCCAGCCTGGGCATTTTCTTCTTGGATGAGGTGCTACCATCTCC  
CAGGGAAGCCACTGAACCCCCAAGGCCCTTCTCCATTTTCTGGCTAAGATAGGACATGGC  
CCATGGACTTTTGAACAACCCAGAGGGGGAACAGCAGTGAATTTCTGGGGAACCCAGGC

15108 TGTACTCATGGTACCCTGGGGGCCCCGCTTCCCACCTGACAGGCAAAGACAGAAAGTG  
TCTGGGAACACTGCCTGGTGGCCGCTGGGCATTTTCTTCTTTTTTCTTTTTCTTTTT  
AGAGATGGAATTTTGCTCTTGTACCCAGGCTTGAGTGCAATGGCGTTATCTTGGCTCAC  
TGCAACCTCCACCTCTGGGGTTCAAGCGATTCTCCTGCCTTAGCCTCCCAAGTCGCTGAG  
ATTACAGGTGCCACCACACCAGCTAATTTTGTATTTTGTAGATATTGGGTTTCACC  
[A, C]  
TGTTGGCCAGGCTGGTGTCAAACCTCCTGACCTCAGGTGATCCACCTACCTTAGCCTTCCA  
AAGTGCTGGGATTACAAGCCTGAGCCACTGCGCCAGCCTGGGCATTTTCTTCTTGGAT  
GAGGTGCTACCATCTCCCAGGGAAGCCACTGAACCCCCAAGGCCCTTCTCCATTTTCTGG  
CTAAGATAGGACATGGCCCATGGACTTTTGAACAACCCAGAGGGGGAACAGCAGTGAATT  
TCCTGGGGAACCCAGGCAGCCAGGGCTAGCAAGGCTGGGGTGGCCATGGCAGTAATCCT

16274 CTTCCAGACTGTCTCCTGTCAGGAGCTGCAGCAGGACTTTGAGAATGCGCCCCCACCGA  
CCCCAACAAACACCAGACCCCGGCTAACGGCACCAGCGTGTCTATATCACCTTCAGCCC  
TGACAGCTCCTCACCTGCCAGAGTGAGCCACCAGCCTCCGCTGAGGCCCCCGGCGAGCC  
CAGTGACATGCTGGCCAGCGTCCCACCTTCGTACCTTCCACACCCTCATCCTGGACAT  
GAGTGGAGTCAGCTTCGTGGACTTGATGGGCATCAAGGCCCTGGCCAAGGTGAGGCCCTC  
[-, G]  
GGGACAGCAAGCACCACCCACTCCACCCCTCCGCTCTGCTCTCCACATTCCCTTTCTCTG  
GGAGCCCTCATTTCAGGAAGCTGAGGGAGGAAGCTCACTGGGGAGACTAACAGCTCCTAG  
GAATCCCTCCTTTCCCAGACGCCACCAGGTTGAGACATTCTCCACAGAGCAGGCCCAGA  
CGGCCCATGACAATGAGTGGCGGGACAAGTCTACCAGAGTTTCAGGCCCTGTGCTCCCA  
ACACCCCCAGCAGTGCCCATCCCAAGTCCCTCTCAGCCATCAGGAACCCACCCAGGTTCT

17424 AACATGGTGAAACCCCGCTCTACTAAAAATACAAAAATTAGCCAGGTGTGGTGACGGGC  
CCCTGTAGTCCCAGCTACTCGGTAGGCTGAGGCAGAGAATTGCTTGAACCCAGGAGCGG  
AGGTTGTCAGTGAGCCAAGATCGCGCACTGCACTCCAGCCTGGGCAACAGAGTGAACCTC  
CATCTCAAAAGAAAAAAAAGAAAATATCTAGCCCCACAAGAAGGGGCCATGGTGACTTT  
AAGTGCCCGCCACGTTGGCAAAAGTCCATTTCCGCTCCACTCCCAGAGAAACCGTCAGC  
[C, T]  
AACACTCCAGGGAGAAGTGGTGTGCTTTGCTGCTATTTTTGTCTTTGGCTGCTGGGCTCT  
CAGGGTTGCTTATTTGTTTGGCTTCCCCTCTGAAGTACGTTTTGTGAATCACTTTTGAGA  
CCCACTCAGAACATTCCCTTTTGCCTCCCTACCCCAACAACACTTCTAGCTGAGCT  
CCACCTATGGGAAGATCGGCGTGAAGGTCTTCTTGGTGAACATCCATGGTAAGAGAAAGA  
GGACATTTAGGGACTGAAAGACTGGCAAGGAGTGTGGGGTAGGAACAGGTTGGTGGGGTC

17627 AATATCTAGCCCCACAAGAAGGGGCCATGGTGACTTTAAGTGCCCGCCACGTTGGCAAAA  
GTCCATTTCCGCTCCACTTCCCAGAGAAACCGTCAGCCAACACTCCAGGGAGAAGTGGTG  
TGCTTTGCTGCTATTTTGTCTTTGGCTGCTGGGCTCTCAGGGTTGCTTATTTGTTTGGC  
TTCCCTCTGAAGTACGTTTTGTGAATCACTTTTGAGCCCACTCAGAACATTCTTTCC  
TTTTGCCTCCCTACCCCAACAACACTTCTAGCTGAGCTCCACCTATGGGAAGATCGGCGT  
[G, A]  
AAGGTCTTCTTGGTGAACATCCATGGTAAGAGAAAGAGGACATTTAGGGACTGAAAGACT  
GGCAAGGAGTGTGGGGTAGGAACAGGTTGGTGGGGTCTGAATAGTGAGGAGGTTGGAAAC  
GAGAGCACCAGCTATCCCCACAAGCTGCTGCCTGCTCATAAAAGCTTCAGGTACAAGT



19571 AGTTTCTGTTGCACAAACACCCAAAGCATAGTAGAGCAGGCCACTCTGCTGGCATCG  
TTCCCTGCCTCCTCCTCATCTCTTTCTAAAGGGGGCTTTGCGGAAGGGAGGGGAGGGGAG  
TAAGCCTACCCATTTAACTTACCGGAGCTTAGAGATTTAGGCTGGTGAGGGATAAAGA  
GATTGGGTCTGAGTTTGTCTCAGCTTTTGACATTTAATTTACTAGCTCAGTAAGTCAT  
ACAAATGGGATACAAATAACACCATCTAAAACTCCAGAAGACTGGGGAGTCAGAAAAATC  
[T, G, C]  
TACCTCCTTGGGGTCCCTGCCCAGATCCCCAGTCATCTCTAGCCCTCAGGGTCCCCCTCCC  
AGCTCAGCTCCTGCCCTTGGCCTCCCAAGACTCTTGTGTGCCCCAGCCCTGGGTAAAAA  
CCTCCCCTGCCCTCTGTGGGTCTAAGAAAGGCTTTCTGGCCCTAGAGCAATGATTTGC  
TCTTTGCCCTTAAGAGACTGATGAAGGTGAAACCATCTGTTCTAAGTGCTGAAAGACTGCC  
CAGGAACACACAGGGCGCTGGCTCCTGCCCTCCATGCCTAGAGGGAAACCCTGGGGAAAC

20147 GCCTAGAGGGAAACCCTGGGGAAACAACGGGCTTTCTGCTTCGTGAAATTTGTCCGCAG  
AGCAAAGAGGGAGATTCTGGAGGAAGCTGCATTAGTTGTTAGTGCCCTAATCATGTTTCAG  
CTACTCTAGTTGGTATGTATACCTGATTAGTCATAGCACTTATAAATAATTTATATTTTA  
TATAATATATACTTACATATTATAGACCATTACAGATACAAATCACACACATAAACACA  
CACCTTTTCAACAGCATTGTGAGGGACAAAGCAGGCAAAGTGAGGCTGGTTATCAGACTT  
[T, G]  
AACAGATTAGAAAATATATTCCCAGGAGGACAGGAATTCCCCAAGGTCAGGCAGCTAGCC  
AATAGTTTTTCTAAGCTGAGTAAACCTTCCCTGCCTCTAACGGCCCAAAAGGAGGGAA  
GACCGCATACACACCTGTCTGGTATAAGGGGGAAGACCACAGCCGTGCTGTTTTGTGA  
GGCAGGTAAAGGGAAGGGGCAAGAGGATAAGTCATGTGTAGGAAGCAGCGTCCAACCAGA  
GCCGGCCACCTGTCCCTTTTCTGCCACCATGCACCAACTTTGCTGTTTCAGTCACTGAAG

20180 TTTCTGCTTCGTGAAATTTGTCCGCAGAGCAAAGAGGGAGATTCTGGAGGAAGCTGCATT  
AGTTGTTAGTGCCCTAATCATGTTTCAGCTACTCTAGTTGGTATGTATACTTGATTAGTCA  
TAGCACTTATAAATAATTTATATTTTATATAATATATACTTACATATTATAGACCATTCA  
CAGATACAAATCACACACATAAACACACACCTTTTCAACAGCATTGTGAGGGACAAAGCA  
GGCAAAGTGAGGCTGGTTATCAGACTTTAACAGATTAGAAAATATATTCCCAGGAGGACA  
[G, A]  
GAATCCCCAAGGTCAGGCAGCTAGCCAATAGTTTTTCTAAGCTGAGTAAACCTTCCCT  
GCCTCTAACGGCCCAAAAGGAGGGAAGACCGCGATACACACCTGTCTGGTATAAGGGGG  
AAGACCACAGCCGTGCTGTTTTTGTGAGGCAGGTAAAGGAAGGGGCAAGAGGATAAGTCA  
TGTGTAGGAAGCAGCGTCCAACCAGAGCCGGCCACCTGTCCCTTTTCTGCCACCATGC  
ACCAACTTTGCTGTTTCAGTCACTGAAGCTCATTCTGCACTGGCTTCCCTCCCTCCAGGCT

20584 TGTCTGGTATAAGGGGGAAGACCACAGCCGTGCTGTTTTTGTGAGGCAGGTAAGGGAAGG  
GGCAAGAGGATAAGTCATGTGTAGGAAGCAGCGTCCAACCAGAGCCGGCCACCTGTCCC  
TTTTCTGCCACCATGCACCAACTTTGCTGTTTCAGTCACTGAAGCTCATTCTGCACTGGC  
TTCTCCCTTCCAGGCTCCAGGGGATGCTGAGCTCTCCTTGTACGACTCAGAGGAGGACA  
TTCGCAGCTACTGGGACTTAGAGCAGGTGAGCTGAGGGAAGGGGCTGTGAGGGTGGGAGC  
[A, T]  
GGGCGAAGAGGGGAAGGATGGGGTGCCTGTCAAATACAAGGCGTTCACTCAGCTGTCTCA  
CCTCCAGCCCAGAGCAGTCACATTCAAGGCCACAAAGATTTGTGGTCATCTTTGTTTTTT  
TTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT  
AGACTGGAATGCAGTGGCATGATCTCAGCTCACTGCAACCTCTGCCTCCCGGGTCCAGA  
GGTTCTCCTGCCTCAGCCTCCCGAGTAGCTGGGACTTCAGGCCTGCGCCAGCTAATTTT

20717 ATGCACCAACTTTGCTGTTTCAGTCACTGAAGCTCATTCTGCACTGGCTTCCCTCCCTTCCA  
GGCTCCAGGGGATGCTGAGCTCTCCTTGTACGACTCAGAGGAGGACATTTCGAGCTACTG  
GGACTTAGAGCAGGTGAGCTGAGGGAAGGGGCTGTGAGGTTGGGAGCAGGGCGAAGAGGG  
GAAGGATGGGGTGCCTGTCAAATACAAGGCGTTCACTCAGCTGTCTCACCTCCAGCCAG  
AGCAGTCACATTCAAGGCCACAAAGATTTGTGGTCATCTTTGTTTTTTTTCTTTTCTTTT  
[T, C]  
CTTTTTTTTTTTTTTTTAAATTTGAGACAAAGTCTCACTCTATCACCCAGACTGGAATGCA  
GTGGCATGATCTCAGCTCACTGCAACCTCTGCCTCCCGGGTCCAGAGGTTCTCCTGCCT  
CAGCCTCCCGAGTAGCTGGGACTTCAGGCCTGCGCCAGCTAATTTTGTATTTTGTAGTA  
GAGACAGCTTTTACCATGTTGGCTGGGCTGGTCTCGAACTTCCGATCTCAAGCAATCTG  
CCTGCCTCGGTCTCCTAAGTGCCTGGATTACAGGCATAAGCCACGATGCCTGGCCTTTGT

20894 GGGGAAGGATGGGGTGCCTGTCAAATACAAGGCGTTCACTCAGCTGTCTCACCTCCAGCC  
CAGAGCAGTCACATTTCAAGGCCACAAAGATTTGTGGTCATCTTTGTTTTTTTTCTTTTCC  
TTTTCTTTTTTTTTTTTTTTTAAATTTGAGACAAAGTCTCACTCTATCACCCAGACTGGAA

09749569-123800



# 034488 - J. Paul Jones

22264 CTCCATGGTGCTCTGAATCCGTCTCTCACAGTGCTGATGGCAGGTGGGACAGATTAGAAA  
ATAGAGCTGGAGCCACAGAGATTGGCAGACTGATTTCCGTGCCCTCTTGAATCTCCAG  
CACATTCCAAAAGCCTGGATAGGACCAAAATAGCTTATCAACGTGAGAAAGGACTTCAG  
AGCTTGCTCTACTGCCAACCCTCATTTTACCCAATGAGGAAAGTGAAGCTATTAGGGGGCG  
AGGGACACGTGGAAGGTCACACAGCACACAGGAGGTGATTCACATGTAGATTTCAGCAC  
[T, C]  
GCTCCTGCCACGCTGGACTGGTTACCTCCTAGGCTGACCCTGCCTCTCCCCTGTTTACA  
CACACTCTCGCACACACACACACACACACACACAGGTGCTTTGTTCTGGCCAGG  
GGTTCTTAGGGTCACTCTTGTTGTCAGCCACTGTGACCCCACTGGTCTAACCTCTCTC  
TTCCCCTCCCACTTCTTCTCTGGTTCTGTCAGGAGATGTTCCGGAGCATGTTTCACGC  
AGAGACCCTGACCGCCCTGTGAGGGCTCAGCCAGTCCTCATGCTGCCTACAGAGTGCCTG

23363 CAGGGACCATGTGCTCTCCACACCCAGGAGTCTAGGCCCTTGGTAACTATGCGCCCCCGT  
CCATCATCCCCAAGGCTGCCCAAACCACCACTGCTGTGAGCAAGCACATCAGACTCTAGC  
CTGGACAGTGGCCAGGACCGTCGAGACCACAGAGCTACCTCCCCGGGGACAGCCCACTA  
AGGTTCTGCCTCAGCCTCCTGAAACATCACTGCCCTCAGAGGCTGCTCCCTTCCCTTGA  
GGCTGGCTAGAAACCCCAAGAGGGGGATGGGTAGCTGGCAGAATCATCTGGCATCCTAG  
[T, C]  
AATAGATACCAGTTATTCTGCACAAAACCTTTTGGGAATTCTCTTTGCACCCAGAGACTC  
AGAGGGGAAGAGGGTGCTAGTACCAACACAGGGAAAACGGATGGGACCTGGGCCAGACA  
GTCCCCCTTGACCCAGGGCCCATCAGGGAAATGCCTTCCCTTTGGTAAATCTGCCTTATC  
CTTCTTTACCTGGCAAAGAGCCAATCATGTTAACTCTTCTTATCAGCCTGTGGGCCAGA  
GACACAATGGGGTCTTCTGTAGGCAAAGGTGGAAGTCTCTCCAGGGATCCGCTACATCCC

FIGURE 3, page 19 of 20

24210

0649E88 31 Feb 00